Temporal integration & healthy ageing

Saija, Jefta Daniël

DOI:
10.33612/diss.97455149

Publication date:
2019

TEMPORAL INTEGRATION

&

HEALTHY AGEING

J. D. Saija

This dissertation was financially supported by:

Rijksuniversiteit Groningen (RUG)
Faculty of Science and Engineering
Heymans Institute for Psychological Research
Universitair Medisch Centrum Groningen (UMCG)
School of Behavioural and Cognitive Neuroscience (BCN)
Netherlands Organization for Scientific Research (NWO)
Netherlands Organization for Health Research and Development (ZonMw)
Rosalind Franklin Fellowship
The Heinsius Houbolt Foundation
Prof. Dr. Eelco Huizinga Stichting

Cover page art: starline / www.freepik.com
Printed by: Ipskamp Printing

ISBN: 978-94-034-2003-5 (printed version)
ISBN: 978-94-034-2002-8 (electronic version).

university of
groningen

Temporal integration
&
healthy ageing

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Monday 14 October 2019 at 14.30 hours

by

Jefta Daniël Saija

born on 9 May 1988
in Assen

Supervisor

Prof. D. Başkent

Co-supervisors

Dr. E.G. Akyürek
Dr. T.C. Andringa

Assessment Committee

Prof. S. Denham
Prof. M.M. Lorist
Prof. R.A. Hut

INDEX

CHAPTER 1

General introduction

In our daily lives, we are constantly trying to make sense of a dynamic, changing world. We receive an enormous amount of input signals through our senses, and our brain has to integrate all this richness so that we are able to understand and interact with our environment. One of the challenges we are faced with in this process is to integrate the information from each sensory modality, allowing us to perceive and identify objects. Consequently, we are able to focus on and pay attention to meaningful objects and events in our surroundings.

Even within a single modality we need to integrate information and use it to identify objects. For example, we are able to perceive a moving, blocky shape on wheels and identify it as a car, or we can hear a typical car engine roaring and infer that a car must be passing nearby. In the visual example, the object's components such as shape, depth and color can be grouped, and after figure-ground segregation has taken place, we match our perception with our representation of a "car" in memory (Wagemans et al., 2012).

In this complex process, time plays an important role, as all objects are subject to the dimension of time. For instance, take the standard frame rate for movies. On average this is about 24 frames per second. If we would watch a movie in which a car drives from the left to the right, while the camera stays stationary, then (depending on the speed of the car) we would see the car driving smoothly. Even though we are presented with 24 individual pictures per second, we still perceive that the car is in motion. Our brain is marvelously able to resolve that the car is the same moving object in each picture. So how is it that we are able to perceive the world fluently and lock representations to recognized objects in time? One of the mechanisms that help us to do this is *temporal integration*, which will be the main focus of this thesis.

TEMPORAL INTEGRATION

One of the basic building blocks that enable the experience of continuous perception is the *functional moment*, which can be said to be the primary level of temporal integration (Wittmann, 2011; Dorato and Wittmann, 2015; Wittmann, 2016). It is characterized as an interval in which there is no clear temporal relation between successive events. In other words, in a functional moment there is no way to tell when one event ends and another starts. Several functional moments, each with different durations, have been identified. An interstimulus interval (ISI) of around 20 to 60 ms is needed in order to perceive the correct temporal order between two visual, auditory or tactile stimuli of short duration of around 1 to 15 ms (Kanabus et al., 2002; Miyazaki et al., 2006; Szymaszek et al., 2009), indicating a functional moment of around 100 ms at most. Also longer functional moments have been identified, ranging up to 300 ms. For example, the total duration of four auditory or visual stimuli is required to be at least 200 to 300 ms to successfully indicate their correct temporal order (Ulbrich et al., 2009). Also, the illusory McGurk effect is only perceived when the auditory and visual stimuli are maximally desynchronized within 200 ms (van Wassenhove et al., 2007). Successive instances of these functional moments are seamlessly enclosed and integrated to result in so-called experienced moments in which we are consciously aware of the present, but in which we also already perceive continuity and temporal relations between events.

Another functional moment can be identified in research on the attentional blink (AB). The AB refers to the decreased ability to identify a second target (T2) after the first target (T1) has been successfully identified. This phenomenon is studied in Rapid Serial Visual Presentation (RSVP) tasks, in which participants have to identify (usually two) targets among a stream of rapidly presented, successive, distractor stimuli. The more distance between both targets, the smaller the AB is. But remarkably, performance is less harmed when T2 follows T1 in direct succession at Lag 1 (i.e., no distractors between targets), which is referred to as Lag 1 sparing. Early studies found that the number of order reversals was highest at this particular Lag (Chun & Potter, 1995; Hommel & Akyürek, 2005) In other words, the judgment of the temporal order between both targets was worst –suggestive of temporal integration- when both targets were presented closest (Akyürek, Toffanin, & Hommel, 2008). This indicates the existence of an even longer functional moment (i.e., around 200 ms; Akyürek et al., 2012) than when the temporal order of two successive stimuli that are presented in isolation have to be determined (i.e., 100 ms at most; Kanabus et al., 2002; Miyazaki et al., 2006; Szymaszek et al., 2009). One major difference between these two is that with the first, the ISI is around 10 ms, which is shorter than that of the latter, which is around 20 to 60 ms. A shorter ISI makes it harder to judge the temporal order between stimuli (Miyazaki et al., 2006), however, it is conceivable that longer stimulus durations will make up for this. Hence, this might be of influence on the difference in duration between functional moments.

In a recent study, Akyurek et al. (2012) investigated the source of increased temporal order reversals of targets at Lag 1. The authors used new sets of visual target stimuli, which had the novel trait of being able to be perceptually combined into new integrated or overlaid targets. For example, if T1 is / and T2 is \, then the integrated target would be X. These integrated targets were valid target responses as well as their individual target components. Importantly, participants were also able to give an empty response, which means that they did not perceive a target. Most trials contained two targets. In these two-target trials, when participants reported seeing only one target, which was the combination of T1 and T2, then it could be concluded that both targets were regarded as a single entity in which temporal information about the individual targets/events was missing.[1] The nature of the target stimuli, together with the short distance between them, should reveal a loss of temporal information about the number of target stimuli and their order, and enable observers to report a new combined percept instead.

[1] Note that participants are not aware that temporal integration of targets is measured, instead they are instructed to identify all separate targets, which puts the focus on segregation instead of integration. So temporal integration is not biased by the instructions or task.

The authors reasoned that in general, the closer two visual events are and the higher their compatibility is, the more likely both events will be perceptually combined into a single entity. In other words, targets in RSVP will be more likely to fall in the same temporal integration window or functional moment at short inter-target lags, up to about 200 ms. The authors found indeed that when this type of compatible stimuli were used, temporal integration reports were high at Lag 1, but order reversals low. But, when they used regular stimuli that were not suitable for integration, the frequency of order reversals jumped back up to typical levels. This type of temporal integration in which events are perceptually combined into a single entity is what we will be referring to as merging temporal integration in the remainder of this introduction and the general discussion section.

THEORIES ON TEMPORAL INTEGRATION

One of the earliest theories on the workings of merging temporal integration that have been proposed is that it is achieved by a sensory storage buffer, which fills up rapidly at stimulus onset and decays at stimulus offset (Coltheart, 1980). At first, this theory has been used to describe *iconic memory* (Sperling, 1960), which refers to a storage of visual information that has a large capacity and fills up rapidly, but afterwards also decays rapidly. Before all items are gone from this storage, a subset can be transferred to a more durable storage that decays less rapidly. Later, this sensory storage buffer was used to describe *visible persistence*, which refers to perceiving a stimulus longer than it is actually physically present (Coltheart, 1980). In other words, the perception of a stimulus persists after stimulus offset. In the *storage hypothesis* (Di Lollo, Hogben and Dixon, 1994), the discharge of items in the storage buffer is hypothesized to result in perceiving stimuli longer than they are physically present. In the case of merging temporal integration, two successive items that are separated by an ISI can be perceptually combined or temporally integrated when the representation of the first stimulus in sensory storage bridges the ISI and (partially) overlaps with that of the second stimulus through visible persistence. The storage hypothesis for visible persistence, however, has not been supported as it was shown that brief stimuli have inverse duration and inverse intensity effects. To clarify, it was found that stimuli with shorter durations and/or lower luminance resulted in longer visible persistence. Theoretically, it is neither logical nor plausible that a shorter and weaker exposure builds up a stronger representation (Coltheart, 1980; Di Lollo, Hogben and Dixon, 1994). Therefore, such a sensory storage buffer is incompatible with visible persistence, and therefore also with merging temporal integration.

Another theory on visible persistence, which is able to explain the inverse duration effect, is the *processing hypothesis* (Di Lollo, 1980). Here, visible persistence is interpreted as a duration of neural activity. Unlike in the storage hypothesis, where a stimulus is said to persist after stimulus offset, in the processing hypothesis persistence starts from stimulus onset and the persistence duration is fixed (+/- 130 ms). Consequently, for stimuli shorter than +/- 130 ms, relatively short stimuli will always seem to have longer persistence, and relatively long stimuli shorter persistence. Like in the storage hypothesis, two stimuli are temporally integrated when the activity of the first stimulus (partially) overlaps with that of the second stimulus. This means that the probability of merging temporal integration depends on the stimulus onset asynchrony (SOA; the duration of the onset of the first stimulus until the onset of the second stimulus): the smaller the SOA, the higher the probability that neural activity of both stimuli overlap. This results in a higher probability that stimuli are temporally integrated. This also means that ISI and stimulus duration influence merging temporal integration equally, because a longer ISI needs a shorter stimulus duration (and vice versa) to maintain the same SOA and the same probability of merging temporal integration.

Di Lollo, Hogben and Dixon (1994), however, found that changing ISI influences merging temporal integration more than changing stimulus duration. Namely, decreasing the probability of merging temporal integration requires a smaller increment in ISI than in stimulus duration of the first stimulus. Also, merging temporal integration is less likely when the second stimulus is longer (Dixon and Di Lollo, 1994). This means that persistence alone is not sufficient for merging temporal integration, as a longer second stimulus does not change the overlap of neural persistence between both stimuli. To account for these observations, the *temporal correlation hypothesis* was developed (Di Lollo, Hogben and Dixon; 1994; Dixon and Di Lollo, 1994). This theory is based on the assumption that our visual system is constantly trying to determine whether consecutive visual inputs (e.g., stimuli or events) are coextensive or disjoint. On the one hand, visual perception has to be fluent, continuous and integrated, and on the other hand, it has to detect small, rapid changes. Whether consecutive visual stimuli are integrated or separated depends on the correlation between the items, and is determined by a temporal coding mechanism. Because consecutive visual stimuli can even be temporally integrated when they are non-overlapping, it is proposed that correlation is not calculated on physical stimuli, but rather on their visual responses (i.e., neural activity), which are delayed in the peripheral and central layers of the visual system and can therefore overlap. The temporal coding mechanism calculates correlations on samples of the visual activities within a sliding window of integration. Within this temporal window, new visual activity is continuously added and old activity decays. Merging temporal integration is more likely when correlation between consecutive stimuli is high, which means that their neural activities or patterns are similar over time. A low correlation means that the activities are too dissimilar, and should therefore be segregated. This entails that correlation is not only based on the degree of temporal overlap of neural activities, as is the case in theories on visible persistence, but also on the similarity between activities and their compatibility. Overall, this hypothesis resembles a more advanced version of the older traveling perceptual moment hypothesis (Allport, 1968).

AUDITORY TEMPORAL INTEGRATION

According to the temporal correlation hypothesis, the temporal code that determines whether successive stimuli will be perceived as integrated or segregated is based on correlations of neural activity of mechanisms that are mostly high-level and central, because merging temporal integration is based on more than merely low-level mechanisms as for example visible persistence. For example, it also takes into account that the duration of the second stimulus has an effect on merging temporal integration, as well as stimulus intensity. Additionally, several studies showed that prior knowledge affects merging temporal integration (e.g., Forget, Buiatti & Dehaene, 2010). In other words, the temporal integration mechanism has to analyze input using prior knowledge and sensory evidence to provide output that makes most sense to us as perceivers. Therefore, it is possible that the temporal integration mechanism is universal, amodal and inherent to perception in general.

An interesting analogy can be made between the visual and auditory domain with respect to temporal analysis of perception. The underlying idea of the temporal correlation hypothesis is that, while our visual perception has to be temporally fluent (supporting long temporal integration), we also need to detect small and rapid changes (supporting sensitivity to short temporal changes). A similar concept can be found in speech perception. To wit, our auditory perceptual system has to analyze syllabic, prosodic and slow-envelope information over the course of 100-200 ms, but in the meantime also has to analyze fast, spectral information (i.e., temporal fine structure) over the course of 20-40 ms (Rosen, 1992). In the asymmetric sampling in time hypothesis, Poeppel (2003) proposed that the brain actually analyzes auditory information asymmetrically: the right hemisphere extracts auditory information from a long window (i.e., 150-250 ms), while the left hemisphere from a short temporal integration window (i.e., 20-40 ms). Similarly, in a theory on multimodal temporal complexity reduction, it is hypothesized that such fast oscillations are periods in which information from different brain areas are treated as co-temporal and integrated (Pöppel, 2009), like a functional moment encompassing multiple modalities. In other words, within one period of an oscillation of about 30-40 ms, temporal information that is spatially distributed in the brain is integrated and treated as one coherent unit. Because it seems that such periods are not modality specific, there might be a possibility that merging temporal integration works similar in other modalities as in vision.

Interestingly, using the auditory equivalent of the RSVP, the Rapid Serial Auditory Presentation (RSAP), Soto-Faraco and Spence (2002) showed that Lag 1 sparing can also occur with the auditory AB. Because we know from the previously discussed visual RSVP studies that Lag 1 sparing correlates with the loss of temporal order information of two visual targets, it might be that this is also the case with auditory stimuli. Consequently, this might mean that merging auditory temporal integration might also span over 200 ms. In a similar fashion, the duration of the functional moment for temporal order judgments of stimuli in isolation is shown to be somewhat similar per modality (Kanabus et al., 2002; Miyazaki et al., 2006; Szymaszek et al., 2009). In chapter 2 of this thesis, we investigated whether merging temporal integration behaves similar in audition as it does in vision.

AGING EFFECTS

Studying merging temporal integration is useful to give us insights in the temporal dynamics and limits of perception. For example, a recent study using an RSVP task showed that there are individual differences in the number of integrated target reports, which means that the duration of the temporal integration window differs per person (Willems et al., 2016). Even though there might be numerous factors that influence the length of an individual's temporal integration window, one of the factors is likely to be age and its incidentals. For example, research on temporal order judgment and gap detection tasks showed that older people need longer ISI and stimulus durations to successfully separate or judge the order of two sequential visual or auditory stimuli (Humes, Busey, Craig, & Kewley-Port, 2009; Kolodziejczyk and Szelag, 2008; Ulbrich et al., 2009). Similar evidence was obtained from visual masking and integration tasks (Di Lollo, Arnett, & Kruk (1982). Such tasks show that older people have lower temporal resolution, indicating that the respective functional moments in which temporal order information is lost are longer. This might indicate that they have longer temporal integration windows.

When individuals become older, besides age-related decline in sensory functions (Swenor et al., 2013), they will usually have to deal with negative, cognitive aspects of aging such as memory problems and reductions in processing speed (Park et al., 2002; Salthouse, 2004; Salthouse 2009). The latter, sometimes referred to as cognitive slowing, is commonly reflected in increased reaction times and is thought to impair cognitive functioning as there is less time available to successfully execute cognitive tasks (Salthouse, 1996; Madden and Allen, 2015). Even though these negative effects can trouble older individuals, such as when one keeps misplacing one's mobile phone, becoming older might also have its merits. For instance, is it always better or beneficial to process the world faster? Research showed that older people see less flickering on monitors with lower frame rates than younger people (Misiak, 1951). If older people's temporal integration windows would be longer due to cognitive slowing, then movies might seem smoother for them as well. However, they might also perceive fewer details. Nevertheless, a pupil dilation study showed that temporally integrating successive visual stimuli instead of segregating them results in less mental effort (Wolff et al., 2015). This eases perceptual processing; for example, in the case of two successive visual stimuli, with merging temporal integration only a single perceptual entity (consisting of both stimuli) has to be formed instead of two distinct entities. This is beneficial for older people because they have less cognitive resources (Salthouse, 1996; Salthouse, 2004; Salthouse 2009), which makes temporal integration a viable compensation mechanism. But do older people have longer temporal integration windows in general? Do they integrate sensory information more over longer intervals, and is this effect comparable in vision and audition? We aimed to answer these questions in Chapter 3 of this thesis.

PHONEMIC RESTORATION

As was previously discussed, in speech perception we need a short temporal integration window to analyze fast spectral changes, but a long window to analyze and extract information from prosody and syllables (Rosen, 1992; Poeppel, 2003). Note that these windows don't facilitate merging temporal integration, but forms of integration in which perceptual information is extracted and analyzed but not necessarily combined into a single percept. If older people have longer temporal integration windows, then this would affect speech perception as well. If the relatively short windows become longer with old age, then fewer details will be perceived. However, if the relatively long windows become longer, then more information might be available over longer segments of speech. Interestingly, this might be beneficial for older people in situations where for example you try to understand someone at a noisy party, where speech is alternately interrupted by noise from other talkers and the environment. This might enable them to better connect the audible speech segments and bridge the inaudible, noisy segments.

Another advantage of old age is that on average the vocabulary and linguistic skills are resistant to age-related decline, perhaps due to experience with language, and stay the same or sometimes even increase with old age (Park et al., 2002; Salthouse, 2004). In challenging listening situations like the noisy party as described above, speech redundancy is reduced and a lot of information might not be accessible, which forces listeners to rely more on their cognitive and linguistic capacities (Stenfelt and Rönnberg 2009). Therefore, having longer windows would enable older individuals to use their expectancies and linguistic skills better to infer what is being said (Pichora-Fuller, 2008). We aimed to investigate this hypothesis using the *phonemic restoration* paradigm in Chapter 4 of this thesis.

In the phonemic restoration paradigm, participants are presented with two conditions, namely speech sentences that have parts of their signal removed at periodic intervals, and with speech sentences in which these parts are filled with loud speech-shaped noise (Warren, 1970). After each speech sentence, they have to repeat what was being said. Due to the degree of uncertainty, participants are encouraged to guess what is the most plausible sentence. This requires the participants to use their linguistic skills and vocabulary to restore the speech stream from individual speech segments, by connecting them through inferring what was said in the noisy segments (Benard, Mensink and Başkent, 2014). The noisy segments remove the spurious cues and spectral splatter at the beginning and end of each suddenly interrupted speech segment. This gives the illusion that the speech continues behind the noise, which activates the phonemic restoration mechanism. This enables the listener to use their vocabulary, linguistic and language skills to infer what was being said during the noise and successfully restore the speech signal. By comparing speech intelligibility between both conditions, we can measure the restoration benefit, or increase in intelligibility, that is obtained by the addition of the speech-shaped noise. Importantly, we can measure whether older people use longer temporal integration windows by comparing the restoration benefit of speech with different speech rates: for older people, because of cognitive slowing, slow speech might be more beneficial as it gives them more time to analyze, and fast speech more troublesome as it makes it harder to keep up with the pace. In fact, longer windows might help older people to use their linguistic knowledge better to solve the speech puzzle when they are given enough time for processing.

OUTLINE

To summarize, in chapter 2, we will discuss whether merging temporal integration works similarly in audition as it does in vision. The knowledge that we gained and the new RSAP task from chapter 2 were used in chapter 3, to investigate whether older people integrate over longer durations in both modalities. In chapter 4, we tested whether older people integrated over longer durations with a measure of (non-merging) temporal integration, in a speech restoration task where successful integration of speech segments enables the better use of linguistic skills. Lastly, in chapter 5 we give a general summary and discussion of all previous chapters.

CHAPTER 2

Temporal integration of consecutive tones into synthetic vowels demonstrates perceptual assembly in audition

J. D. Saija, T. C. Andringa, D. Başkent and E. G. Akyürek

Published in Journal of Experimental Psychology. Human Perception and Performance (2014), 40(2), 857–869. doi: 10.1037/a0035146.

ABSTRACT

Temporal integration is the perceptual process combining sensory stimulation over time into longer percepts that can span over ten times the duration of a minimally detectable stimulus. Particularly in the auditory domain, such “long-term” temporal integration has been characterized as a relatively simple function that acts chiefly to bridge brief input gaps, and which places integrated stimuli on temporal coordinates while preserving their temporal order information. These properties are not observed in visual temporal integration, suggesting they might be modality-specific. The present study challenges that view. Participants were presented with rapid series of successive tone stimuli, in which two separate, deviant target tones were to be identified. Critically, the target tone pair would be perceived as a single synthetic vowel if they were interpreted to be simultaneous. During the task, despite that the targets were always sequential and never actually overlapping, listeners frequently reported hearing just one sound, the synthetic vowel, rather than two successive tones. The results demonstrate that auditory temporal integration, like its visual counterpart, truly assembles a percept from sensory inputs across time, and does not just summate time-ordered (identical) inputs or fill gaps therein. This finding supports the idea that temporal integration is a universal function of the human perceptual system.

INTRODUCTION

Stimulus detection thresholds and stimulus duration are inversely related. In other words, the threshold for detecting an auditory stimulus decreases when its duration increases. For normal hearing listeners, each tenfold in duration corresponds on average to a threshold drop of 8 to 10 dB (Hughes, 1946; Plomp & Bouman, 1959), and this relation holds for stimulus durations of a few hundred ms. When stimulus intensity is held constant (Munson, 1947), the perceived loudness of a tone increases gradually from onset until a steady loudness is reached at a certain duration. These effects are often described as the temporal integration of acoustic energy. It is usually modeled as a leaky integrator (cf. Viemeister & Wakefield, 1991) that sums up acoustic energy over time within frequency bands, but leaks energy exponentially (Plomp & Bouman, 1959; Zwislocki, 1960). Various models of temporal integration have been proposed in terms of electric circuits (Jeffress, 1967; Munson, 1947) and neural excitation (Zwislocki, 1960). These models usually assume a relatively long temporal window of about 200 ms, a duration in line with psychophysical observations, which make these models perfect for explaining integration phenomena like threshold reduction and loudness augmentation.

Multiple stimuli in one memory trace

Recent studies on auditory sensory memory have supported the idea that auditory stimuli are integrated over such comparatively long time intervals. In this field, many studies have used electroencephalography to measure a component of the event-related potential called the mismatch-negativity (MMN). The presence of an MMN after stimulus presentation means that a violation of the norm in a series of stimuli is perceived. Any deviation in a to-be-expected order or identity of sequential stimuli can elicit an MMN, including deviations from preceding stimuli that are represented by a short-term memory trace in the auditory cortex (for reviews see Näätänen, Paavilainen, Rinne, & Alho, 2007; Näätänen, Kujala, & Winkler, 2011).

An MMN study by Tervaniemi, Saarinen, Paavilainen, Danilova, and Näätänen (1994), who studied the effect of deviations in tone pairs on the MMN, suggested that two closely spaced stimuli, with an inter-stimulus interval (ISI) of maximally 140 ms, can be integrated into a single unitary sensory event. Yabe et al. (1998) came to a similar conclusion while investigating the effect of stimulus omission in trains of stimuli with different stimulus onset asynchronies (SOAs) on MMN responses obtained with magnetoencephalography, with temporal integration window estimate at around 160 – 170 ms (cf. Yabe et al., 1998), although others have estimated this to be just slightly longer, at around 200 ms (Sussman, Winkler, Ritter, Alho, & Näätänen, 1999).

In their influential review paper, Näätänen and Winkler (1999) concluded that auditory temporal integration is not merely a process of reducing auditory noise by compressing the time dimension (Näätänen, 1995), such as bridging a small gap or summing up energies, but is rather a constructive process which combines auditory information (pitch, loudness, duration, location and energy) into a single perceptual event. This idea is also consistent with the larger concept of auditory scene analysis, a general model of auditory perception where signal components that are produced by the same source are perceptually grouped into auditory objects (Bregman, 1994).

Importantly, Näätänen and Winkler (1999) proposed that an auditory episodic memory trace is established when combined input from different acoustic feature detectors is placed on "temporal coordinates" (i.e., preserving temporal order information within the trace). The authors posited a parallel between the medium of space, which is central to visual feature integration (e.g., Treisman, 1996), and that of time, which is central to auditory integration. Only after this temporal trajectory is established does the memory trace constitute a genuine acoustic object that can be perceived and experienced subjectively. The formation of these object representations is assumed to occur within a continuous sliding temporal integration window of about 200 ms (Näätänen, 1990 as in Näätänen & Winkler, 1999), although the temporal window of integration might also start at stimulus onset (Yu et al., 2011). Either way, this conceptualization of temporal integration in audition seems like a free lunch: Forming an integrated percept while fully preserving all temporal information suggests that temporal integration is costless in terms of maintaining the properties of the input signal. The current study sought to investigate this claim, because there is evidence to the contrary from visual paradigms.

Similarities to vision

Assuming that auditory and visual perception operate on similar principles, studies on visual temporal integration may provide important insights into auditory temporal integration. In the so-called missing element task (MET; Akyürek, Schubö, & Hommel, 2010), observers view stimuli that are arranged in an evenly-spaced square grid, across two successive partial displays (e.g., Hogben & Lollo, 1974). For instance, using a grid of 25 positions (5x5), observers are first shown a set of 12 stimuli, and then another set of 12 (i.e., 24 in total). Observers are asked to locate the one remaining empty position. Finding the missing element is virtually impossible by mentally comparing and examining the two stimulus displays. When those two displays are temporally integrated, however, they appear as if they were overlaid and then the missing element is immediately apparent. Since temporal integration is more likely to occur at shorter SOAs, the typical finding in the MET is that shorter SOAs result in higher task performance. Evidence from the MET shows that although information about individual parts appears to be inaccessible, the sum thereof still is, and constitutes the integrated percept. This contrasts with the findings from previously discussed auditory studies, which suggested that information about individual parts can be accessed while also being combined into an integrated percept.

Further data on the nature of visual temporal integration has been obtained in studies that investigated performance in dual-target rapid serial visual presentation (RSVP) tasks. In such tasks two targets (T1 and T2) of short duration are presented among distractors in rapid succession (often with short blank gaps in-between stimuli), and the participant is asked to report the identity and order of the targets. T2 can follow T1 with or without distractors in between and this distance is denoted as lag. Lag 3 for example means that T2 follows T1 with two distractors in between, thus T2 lags T1 as the third item. In RSVP tasks, participants often fail to report T2 when it follows T1 closely, within +/- 500 ms after T1 onset (Broadbent & Broadbent, 1987; Raymond, Shapiro, & Arnell, 1992); a phenomenon known as the attentional blink (AB). There is one salient exception to the AB: When T2 follows T1 immediately at Lag 1, without distractors in-between, it is often identified quite well. This exception is called the Lag 1 sparing effect.

Further to the special status of Lag 1, Hommel and Akyürek (2005) showed that although the identity of both targets is often retained, their temporal order is often lost; instead of reporting T1 as the first target and T2 as the second, observers frequently report T1 as the second target and T2 as the first. The frequency of these order errors furthermore varies with the expectations of the observers with regard to stimulus presentation speed (Akyürek, Toffanin, & Hommel, 2008). The authors interpreted these order errors as a consequence of the temporal integration of the two targets into one event representation, and concluded that temporal integration is likely to play a dominant role at Lag 1 in RSVP. This was confirmed by Akyürek et al. (2012), who presented target stimuli that formed reportable identities not only when viewed individually, but also when combined. They used targets such as "/" and "\" that could be perceptually combined to form an "X", which itself was then also a possible target identity. In this task observers frequently reported having seen only the integrated percepts at Lag 1 (at the expense of order errors), confirming the expected effect of temporal integration at this lag. Taken together, these RSVP studies thus suggest that although temporal integration may facilitate visual target identification, it does come at a price—information about the sequence of individual stimuli is lost.

In summary, it seems that in vision, like in audition, two stimuli can also be bound to a single memory trace. Yet, an obvious discrepancy also exists. Whereas in vision temporal integration seems to be associated with a loss of temporal order of the stimuli that are part of the integrated percept, auditory studies, in particular those examining the MMN, suggest that such temporal information is mostly retained. Note that it is entirely possible that this apparent difference between modalities exists as a consequence of the different roles of time in vision and audition: One might argue that the importance of time in audition may render it immune to losses that are incurred in vision, in which spatial information may dominate.

Current research

The present study sought to examine the degree to which temporal information and stimulus individuality might be retained in auditory temporal integration, and whether (these aspects of) temporal integration might be modality-specific. More specifically, the study aimed to provide more definitive evidence of how auditory temporal integration works and to investigate what models are most plausible. To this end, an auditory task similar to RSVP was developed, in which temporal integration of two strictly successive target stimuli was likely. In this rapid serial auditory presentation (RSAP) task (see e.g., Horváth & Burgyán, 2011; Tremblay, Vachon, & Jones, 2005), the targets were chosen in such a way that both the successive report of individual targets, as well as their combined report, were possible (similar to Akyürek et al., 2012). Targets of the study consisted of pairs of first and second formants (harmonic complexes bandpass filtered at specific frequencies) and the 2-formant combined synthetic vowels. In other words, participants were able to report hearing an integrated percept of the sequentially presented formants, which would be equal to a simultaneous presentation thereof (i.e., a 2-formant vowel). Reports could thus vary between having heard T1 first and T2 second, T2 and then T1 (order error), or T1+T2 (integration of first and second formants into 2-formant combined synthetic vowels), and any partial version in which either target was missed.

Three versions of the RSAP task were implemented: In Experiment 1, natural differences in formant intensity of the formant pairs as measured from spoken Dutch vowels (Pols, Tromp, & Plomp, 1973) were used for the successive targets. The use of natural differences in intensity means that the first formant (F1) is always of higher intensity than the second formant (F2), resulting in a more natural percept of the 2-formant vowels. However, in the visual domain, a large contrast between the physical properties of T1 and T2 can also have an effect on attentional blink and the sparing effect (Chua, 2005; Experiments 2a and 3, Table 1). Therefore, to rule out any additional effects due to differences in intensity (and the resulting loudness), loudness difference was minimized in Experiment 2, where formants of equal loudness, based on the equal-loudness contour (ISO 226, 2003), were used. As a consequence, the vowels in Experiment 2 sounded less natural, which also provided a measure of the extent to which natural language familiarity might contribute to integration. Finally, Experiment 3 was performed to investigate the possible effects of the response alternatives that were available to the participants. Because the majority (5/7) of response keys in Experiment 1 and 2 represented vowels, this might have induced a general bias towards reporting vowels. Therefore the number of vowel response keys was reduced (to 1/3) in Experiment 3.

The predictions were as follows. If temporal integration in audition retains temporal coordinates, as suggested by previous work, then the integration of the targets in the present task at short lags (i.e., Lag 1) should result in an increase in the number of correct reports, that is, an escape from the attentional blink. However, neither reports of illusory simultaneous percepts, nor the frequency of order errors should be increased. However, if temporal integration in audition behaves similarly to that of in vision, then reports of integrated percepts should be frequent. This would support the idea that temporal integration is a central, modality-unspecific perceptual function.

EXPERIMENT 1

Experiment 1 investigated whether two auditory targets could be integrated and reported as a single integrated percept, using natural intensity differences of the first two formants of naturally spoken Dutch vowels.

Method

Participants

Sixteen (13 female, 3 male) normal hearing (< 20 dB Hearing Level measured at .25, .5, 1, 2, 4, and 6 kHz) and native Dutch-speaker students of Psychology Department at the University of Groningen participated in the experiment for course credit. Mean age was 20 years (range 18-23 years). Participants were unaware of the purpose of the experiment. Informed consent was obtained in writing and ethical approval was obtained from the local ethical committee of the Psychology Department.

Apparatus and stimuli

The experiment was programmed in Matlab (7.10.0.499 32-bit) using Psychtoolbox (3.0.9; Brainard, 1997; Pelli, 1997) and run under Max OS X (10.5.8) on a Mac Pro equipped with a quad-core Xeon CPU and 8 GB RAM. Participants were tested in a sound-isolated booth. Sounds were presented diotically through a Sennheiser HD 600 headphone, connected to an Echo Audiofire 4 external soundcard and a Lavry Engineering DA10 digital-to-analog converter. Responses were collected with a standard keyboard.

Target stimuli consisted of first and second formants (F1 and F2), harmonic complexes bandpass filtered (specifics below, and in Table 1) at the formant frequencies, of the 5 Dutch vowels /a/ (as in h*aa*t), /i/ (as in h*ie*t), /I/ (as in h*i*t), /ø/ (as in h*eu*t) and /y/ (as in h*uu*t). The synthetic vowel that would result from simultaneous presentation of these formant pairs was also a possible target identity so that the participants could illusorily report a vowel, but it was only rarely an actual target (i.e., on some of the single-target trials). A complex tone with a center frequency of 1 kHz, produced with the same bandpass filter as for the formants, was used as a repeating distractor. 1 kHz lies between the F1 and F2 values, and therefore fits well with the task that required participants to identify F1 and F2 as low and high tones, respectively. The vowels were specifically chosen, based on the distance in frequency of both formants to the 1 kHz boundary and on the relative distance of the formants between vowels. Larger frequency distances between formants and the 1 kHz boundary were aimed for to increase discriminability between the five vowels.

The formants and distractor stimuli were created by applying an infinite impulse response (IIR) filter (Carlyon, Deeks, Norris, & Butterfield, 2002; Heinrich, Carlyon, Davis, & Johnsrude, 2008; Rabiner & Schafer, 1978) at the desired center frequency (see Table 1; based on Pols et al., 1973) to a harmonic complex of 120 Hz with 100 harmonics and a sampling rate of 44.1 kHz. The filter orders for /a/, /i/, /I/, /ø/, /y/ and the distractor were 6, 10, 4, 6, 10 and 8, respectively, and were empirically chosen based on achieving a balance between creating tone-like stimuli for single targets and vowel-like stimuli once formants were combined. The 3-dB bandwidth of the filter was set at 90 Hz. F1 was presented at 65 dB sound pressure level (SPL), but F2 was presented at a lower SPL than the F1 of the same vowel, according to the intensity differences between formants observed in natural speech (Pols et al., 1973). The vowels, i.e., combined formants, as well as the distractors were presented at 65 dB SPL. Figure 1 shows spectrograms, which illustrate the formant stimuli (F1 and F2 of the five vowels) in the lower panels, and an example trial of Lag 3 containing the F1 and F2 of the vowel /a/ together with the surrounding distractors in the upper panel.

Table 1. Frequencies of F1 and F2, and deviations of F2 intensity from F1 intensity, in dB SPL (adapted and modified from Pols, Tromp and Plomp, 1973)

	/a/	/i/	/I/	/ø/	/y/
F1 in Hz	795	294	388	443	305
F2 in Hz	1301	2208	2003	1497	1730
Deviation of F2 intensity from F1 intensity (dB SPL)	-5.6	-19.5	-17.3	-15.6	-18.1

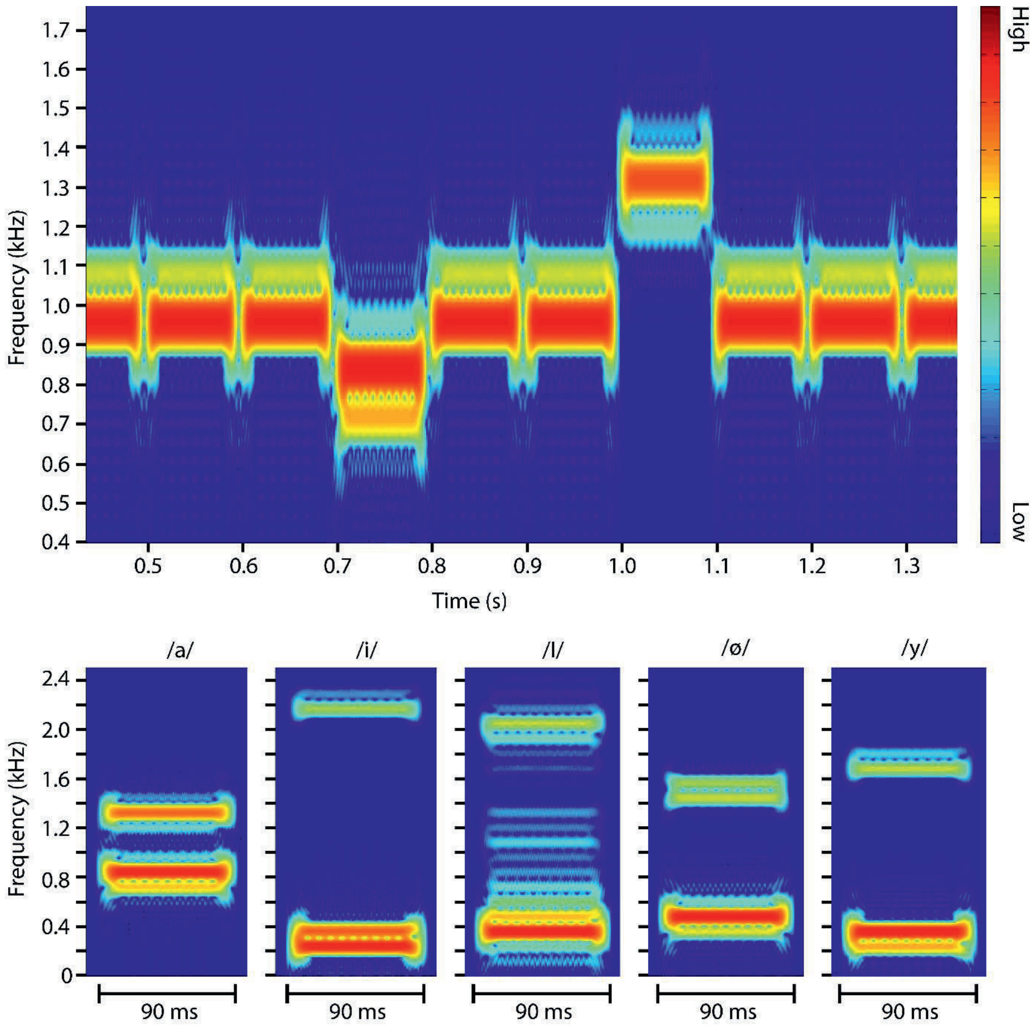


Figure 1. Representation of the stimuli. The top of the figure shows the spectrogram that illustrates a part of a Lag 3 trial. Energy values are represented by different color gradients and range from low (dark blue) to high values (dark red). Complex tones are represented by high concentrations of energy, which last 90 ms and are followed by a silent gap of 10 ms. This example illustrates the mid-section of a Lag 3 trial, where first distractor tones are presented, followed by a low tone (F1 of /a/), then two distractor tones, and a high tone (F2 of /a/) followed by more distractor tones. The five spectrograms at the lower half of the figure illustrate the five 2-formant vowels /a/, /i/, /I/, /ø/ and /y/ that were combined by adding the corresponding first and second formants.

Procedure and design

Participants were unaware that among the stimuli 5 different F1s and F2s were used. Instead they were told that the targets consisted of a random low tone (which was an F1), a random high tone (which was an F2) and five vowels. A low tone was defined as any given F1 tone that was lower in frequency than the distractor and a high tone as any given F2 tone higher than the distractor. All 7 possible targets were labeled on the numerical keypad, so that participants did not have to memorize which target corresponded to which key on the keyboard.

Participants had to be acquainted to the vowels, learn to distinguish them and also learn to classify a low and high tone with respect to the distractor. Therefore, in the first session, participants could press any of the labeled keys to hear a stimulus until they felt they could distinguish all five vowels and knew the difference between a low tone, high tone and distractor. After that session, there was a short training with feedback in which stimuli were presented and participants had to report which of the stimuli they heard. This training session was completed within 15 minutes on average. Once participants successfully learned to distinguish the stimuli, there was a short block of practice trials. The only feedback provided was the playback of the sound of the participant's response, so that the participants could compare their response to what was heard in the trial. After that, the real experiment began which consisted of 605 trials with no feedback. A trial consisted of a stream of 18 consecutive items; in this stream there could be either 1 or 2 targets, the rest of the items were distractors. On 92.6% of all trials there were two targets. In these two-target trials both formants of a particular vowel were required to be targets (i.e., T1 was F1, T2 was F2 or vice versa). T1 could appear as fifth, sixth, seventh or eighth item. T2 followed T1 with 0, 2 or 7 distractors in-between (Lag 1, Lag 3 and Lag 8, respectively, and 39.7%, 26.4% and 26.4% of all trials, respectively). T1 was a solo target in 7.4% of all trials, in which T1 could be a single formant (low tones, 2.47%; high tones, 2.47%) or vowel (2.47%). Each item had a duration of 90 ms, determined in a pilot study, and between the items there was a gap of 10 ms; this gave an SOA of 100 ms. The different conditions are illustrated in Figure

2. Each trial started when the space-key was pressed, and participants could take a break between trials. After each trial the participant was asked to enter what they heard as first and second target in the correct order. If no first or second target was heard, they could press the enter key for an empty response. Reporting only one target without entering a second one could thereby be counted as a solo response. The experiment lasted approximately 60 minutes.

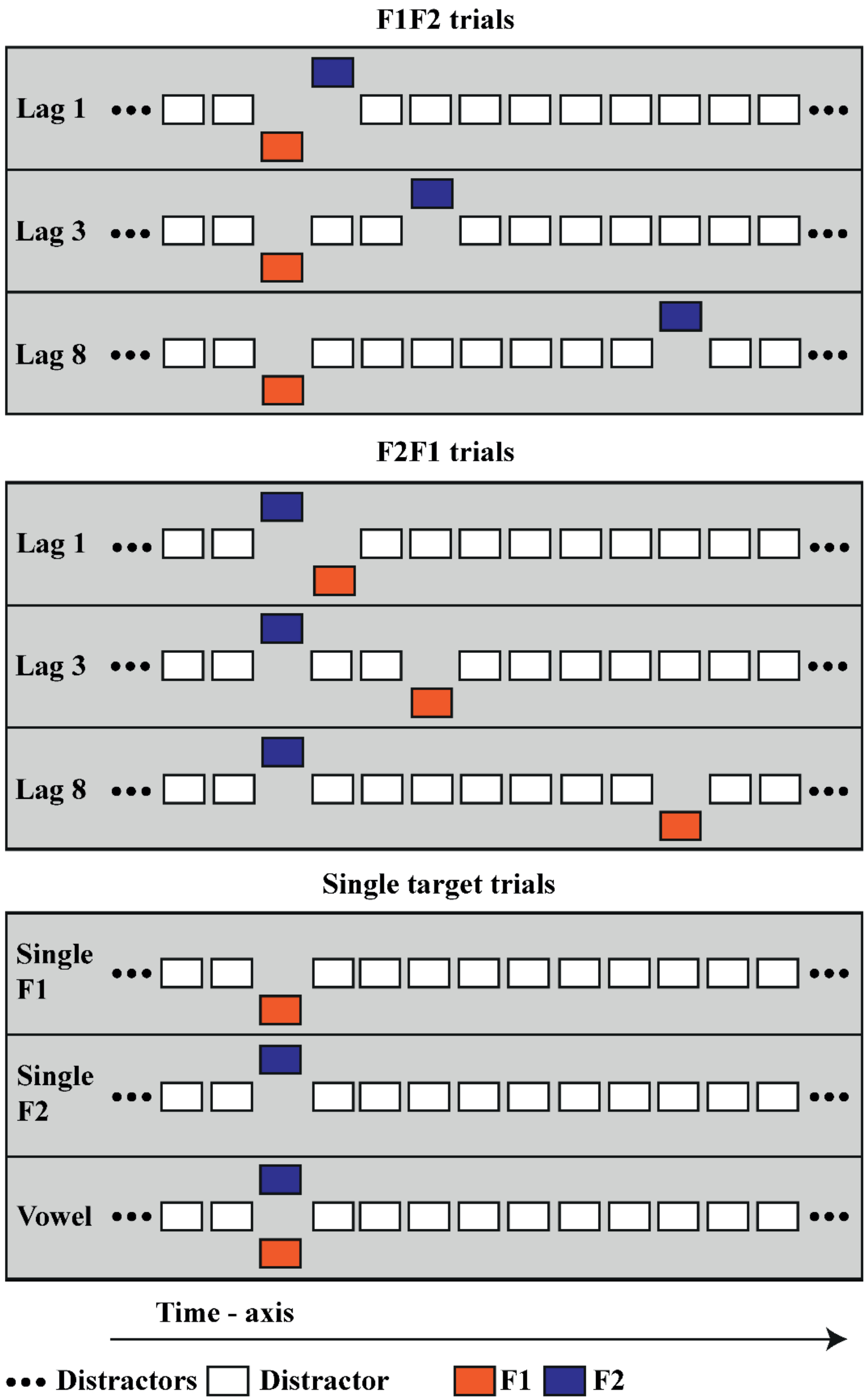

F1F2 trials
Lag 1
Lag 3
Lag 8
F2F1 trials
Lag 1
Lag 3
Lag 8
Single target trials
Single F1
Single F2
Vowel
Time - axis
Distractors
Distractor
F1
F2

Figure 2. Schematic representation of the different conditions. The number that accompanies the Lag indicates the temporal delay between the first and second target, e.g., Lag 3 means that T2 lags T1 as the third successive stimulus with two distractors in-between. The height of the items indicates the relative frequency differences, e.g., F1's have lower frequency than distractors, which in turn have lower frequency than F2's. Targets as well as distractors lasted 90 ms, followed by a silent gap of 10 ms.

Data analysis

First, task performance was examined by analyzing the mean accuracy of T1 and (T2|T1) at Lag 1, 3 and 8. (T2|T1) stands for the accuracy of T2 in cases when T1 was correct. Note that in these analyses a target is only considered correct if both identity and temporal order have been successfully reported. Each analysis consisted of a repeated measures analysis of variance (ANOVA) with the single variable of Lag (1, 3 or 8). In these ANOVA's, when sphericity was not assumed, degrees of freedom were adjusted using the Greenhouse-Geisser epsilon correction. The same analyses were performed for frequency of strict integrations (i.e., only a single integrated response reported) and order reversals (i.e., both targets reported in the incorrect order). Strict integrations and order reversals are cases where both target identities were preserved; these analyses were therefore conducted relative to the total number of trials on which both target identities were preserved. An example of a strict integration response occurs if T1 is F1 (low tone) and T2 is F2 (high tone) of the vowel /I/ and /I/ is given as a solo response. This indicates that both targets (and thus formants) have been integrated into a single representation of the particular vowel and no second target is perceived. Furthermore, to assess the presence of the attentional blink, a paired samples t-test was used to compare T2|T1 identification accuracy at Lag 1 to Lag 8. Additionally, all analyses were performed on rationalized arcsine transformed scores. The statistical outcomes of these transformed scores are reported (in footnotes) when they differed from the analyses on untransformed scores, In all analyses, an alpha level of 0.05 was used. Each analysis is clarified by line or bar graphs. The line graphs that show strict integrations and order reversals together are shown relative to the total number of trials on which both target identities were preserved, while the bar graphs show absolute report frequencies.

Results and discussion

T1 accuracy was strongly affected by Lag, $F(1.4, 20.5) = 17.489$, $MSE = 0.004$, $p < 0.001$. Performance averaged 20.1% at Lag 1, compared to 27.1% at Lag 3, and 31.5% at Lag 8. When report order was ignored performance was 49.2% at Lag 1, 56.6% at Lag 3 and 60.5% at Lag 8. This is illustrated by the left panel of Figure 3.

The accuracy for (T2|T1) was affected by Lag, $F(2, 30) = 5.081$, $MSE = 0.013$, $p < 0.015$. Performance averaged 14.4% at Lag 1, compared to 25% at Lag 3, and 25.7% at Lag 8. A paired samples t-test showed a significant difference between Lag 1 and Lag 8 ($t(15) = -2.989$, $MSE = 0.038$, $p < 0.01$), indicating an early attentional blink (cf. Horváth & Burgyán, 2011; Tremblay et al., 2005). It also indicated, as is often observed in RSAP tasks, that there was no Lag 1 sparing. When report order was ignored, performance was 67.7% at Lag 1, 69.9% at Lag 3 and 71.5% at Lag 8. This is illustrated by the right panel of Figure 3.

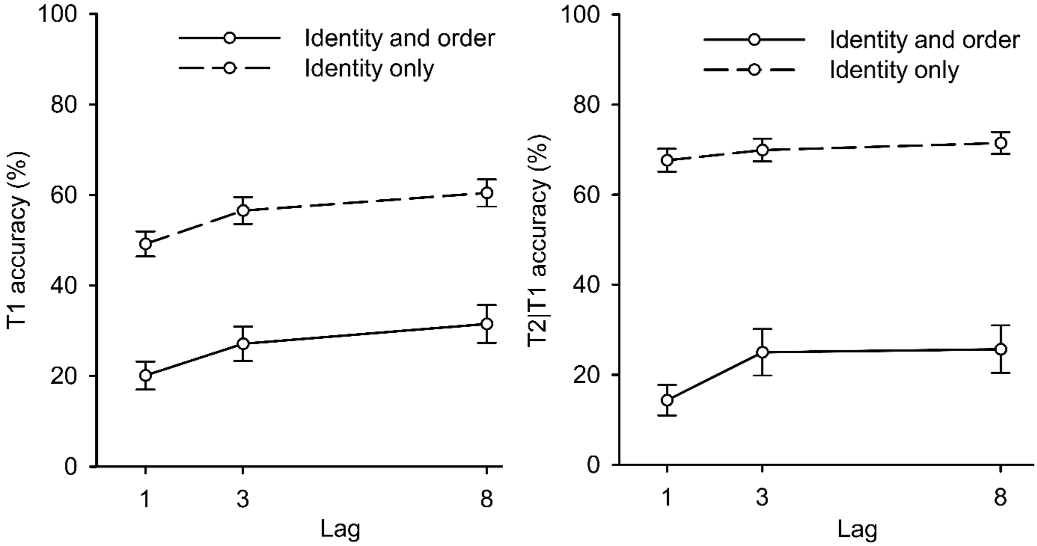


Figure 3. Experiment 1: The left panel shows task performance on T1 in percent correct, plotted over Lag (T2 being first, third or eighth stimulus after T1). Error bars represent ± 1 standard error of the mean. The right panel shows T2 performance given that T1 was correctly reported (T2|T1) in percent correct plotted over Lag. Dashed lines represent identification accuracy if report order is ignored (relaxed accuracy criterion).

Importantly, the frequency of strict integrations was strongly affected by Lag, $F(2, 30) = 20.093$, $MSE = 0.026$, $p < 0.001$. Integrations averaged 66.9% at Lag 1, compared to 41.9% at Lag 3, and 31.8% at Lag 8. Order reversals were not affected by Lag, $F(2,30) = 2.939$, $MSE = 0.008$, $p = 0.068$. Reversals averaged 8.1% at Lag 1, compared to 15.6% at Lag 3, and 11.5% at Lag 8.[1]

[1] Analyses on the rationalized arcsine transformed scores show that order reversals were affected by Lag, F(2, 30) = 4.392, MSE = 117.479, p < 0.05. Reversals averaged 1.6 Rational Arcsine Units (RAU) at Lag 1, compared to 12.8 RAU at Lag 3, and 8.9 RAU at Lag 8.

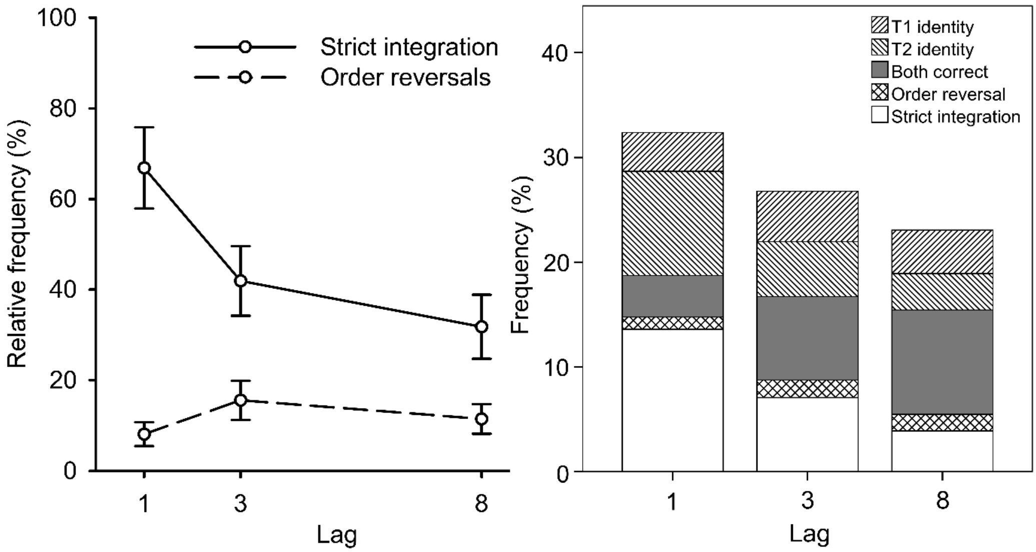


Figure 4. Experiment 1: The left panel shows the relative frequency of strict integrations and order reversals plotted over Lag, as a percentage of the total number of responses in which both target identities were preserved. The right panel shows the distribution of responses for each lag, as a percentage of the total number of responses.

Figure 4 illustrates that the number of strict integrations was higher at Lag 1 compared to later lags. This suggests that two distinct auditory stimuli that succeed each other in a short interval, without actually overlapping or being physically continuous, can indeed be temporally integrated in such a way that a meaningful percept is constructed. The report of such integrated percepts implies that its constituent tones were perceived as if they were simultaneous; a complete loss of order information similar to that observed in visual temporal integration (Akyürek et al., 2012). In this context it is important to note that singular integrations (i.e., without entering a second response) were reported despite deliberate biases in the task towards the report of two individual tones, which were by far the most frequent stimuli, and the most frequent type of trial. Indeed, at later lags, increased reports of the two individual targets were observed. At these lags the succession between targets is too slow and together with the presence of intervening distractors, makes integration unlikely.

EXPERIMENT 2

Experiment 2 was designed to eliminate potential effects of intensity contrast between F1 and F2, as well as possible resultant language familiarity effects, as discussed before, by presenting all stimuli at the same loudness.

Method

Participants. Sixteen (12 female, 4 male) new participants were included using the same procedures and criteria as in Experiment 1. The mean age was 20 years with a range of 18 to 25 years.

Apparatus and stimuli. The experimental setup and stimuli were the same as for Experiment 1. The only difference was that the relative intensity differences between formants from Table 1 were not used. Instead, each stimulus was presented at the same loudness, determined using the equal-loudness contour (ISO 226, 2003). This contour gives estimates of what intensity level in dB SPL is needed in order for a stimulus to sound subjectively equally loud as a stimulus of 1 kHz at a particular loudness level in phons. Table 2 shows the values in dB SPL that were obtained by the calculations using the equal-loudness contours. All F2s were adjusted to these values. Vowels were presented at the average sound pressure level of both corresponding formants.

Table 2. Sound pressure levels calculated with equal loudness contours (with reference to 65 phon at 1 kHz, shown in the table as distractor)

	DISTRACTOR	/a/	/i/	/I/	/ø/	/y/
F1 center frequency (Hz)	1000	795	294	388	443	305
F1 intensity (dB SPL)	65	64.8	70.3	68.2	67.4	70
F2 center frequency (Hz)	1000	1301	2208	2003	1497	1730
F2 intensity (dB SPL)	65	67.7	63.6	65.1	68.5	67.6

Procedure and design

The procedure and design were the same as in the previous experiment.

Results and discussion

T1 accuracy was not affected by Lag, $F(1.3, 19.1) = 3.441$, $MSE = 0.007$, $p = 0.071$. Performance averaged at 26.6% at Lag 1, compared to 31.3% at Lag 3, and 32.2% at Lag 8. When report order was ignored performance was 54.9% at Lag 1, 61% at Lag 3 and 61.1% at Lag 8. This is illustrated in the left panel of Figure 5.

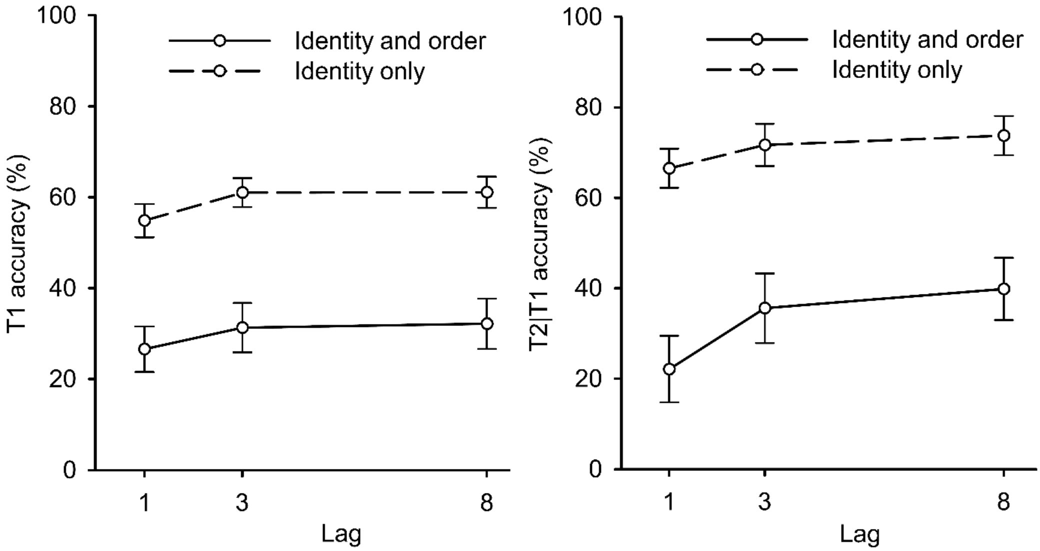


Figure 5. Experiment 2: The left panel shows T1 task performance for each lag. Error bars represent ± 1 standard error of the mean. The right panel shows (T2|T1) performance for each lag. Dashed lines represent identification accuracy if report order is ignored.

Accuracy for (T2|T1) was strongly affected by Lag, $F(2,30) = 10.006$, $MSE = 0.014$, $p < 0.001$. Performance averaged at 22.1% at Lag 1, compared to 35.6% at Lag 3, and 39.8% at Lag 8. A paired samples t-test showed a significant difference between Lag 1 and Lag 8 ($t(15) = -3.896$, $MSE = 0.045$, $p < 0.001$), indicating the expected early attentional blink, similar to the previous experiment, despite using equal loudness for all stimuli. When report order was ignored, performance was 66.5% at Lag 1 , 71.7% at Lag 3 and 73.8% at Lag 8. This is illustrated in the right panel of Figure 5.

The frequency of strict integration was again strongly affected by Lag, $F(1.3, 19.4) = 23.280$, $MSE = 0.058$, $p < 0.001$. Integrations averaged 60.9% at Lag 1, compared to 20.9% at Lag 3, and 19.7% at Lag 8. Order reversals were not affected by Lag, $F(1.3, 19) = 2.297$, $MSE = 0.036$, $p = 0.142$. Reversals averaged 7.2% at Lag 1, compared to 17.8% at Lag 3, and 8.6% at Lag 8.[2]

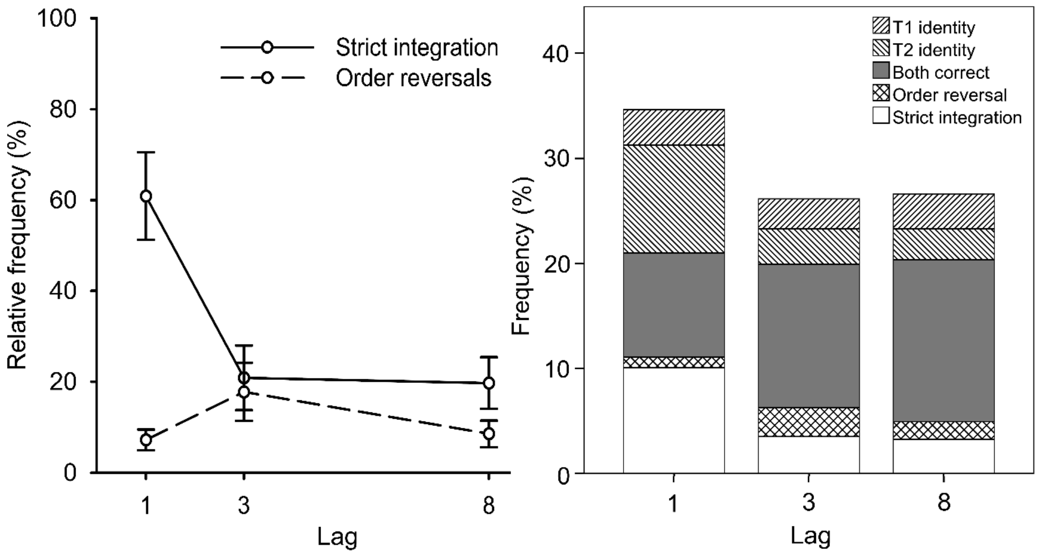


Figure 6. Experiment 2: The left panel shows the relative frequency of strict integrations and order reversals for each lag, as a percentage of the total number of responses in which both target identities were preserved. The right panel shows the distribution of responses for each lag, as a percentage of the total number of responses.

Figure 6 also illustrates how, similar to Experiment 1, the relatively high number of integrations at Lag 1 stands in contrast to that at the longer lags. The number of order reversals was not affected by Lag and seemed, similar to Experiment 1, unrelated to integration frequency. Overall, Experiment 2 replicated the results of Experiment 1. It can thus be concluded that temporal integration was not the result of the loudness differences

[2] Analyses on the rationalized arcsine transformed scores show that order reversals were affected by Lag, $F(2, 30) = 3.472$, $MSE = 222.842$, $p < 0.05$. Reversals averaged 1.7 RAU at Lag 1, compared to 14.3 RAU at Lag 3, and 2.8 RAU at Lag 8.

between the stimuli that were used in Experiment 1, and was thus also unlikely to result from the degree of familiarity with the vowels used in the task.

EXPERIMENT 3

Experiment 3 was conducted to eliminate a possible response bias towards the report of vowels by reducing the number of vowel response alternatives. To this end, the number of vowel stimuli (and consequently the respective F1s and F2s) was reduced from five to three. Next to these three vowel response alternatives, participants now had the opportunity to identify the six remaining tones (rather than just classify as high or low), which thereby made up the majority of the response alternatives (6/9).

Method

Participants. Fifteen (9 female, 6 male) normal hearing (< 20 dB Hearing Level measured at .25, .5, 1, 2, 4, and 6 kHz) and native Dutch-speaker students of Psychology Department at the University of Groningen participated in the experiment following the same procedure as in Experiment 1. Mean age was 21 years (range 20-23 years).

Apparatus and stimuli.

Apparatus and stimuli were similar to that of Experiment 2, except that only three Dutch vowels were used as stimuli: /a/ (as in h*aa*t), /i/ (as in h*ie*t) and /ø/ (as in h*eu*t).

Procedure and design. The task differed from the previous two experiments such that when a tone was heard as a target, the participants not only had to classify it as low or high with respect to the filler tone, but they additionally had to identify the correct tone among three different low and three different high tone options. Thus, the response alternatives were three vowels, three low and three high tones. This increased task difficulty, but more importantly removed any response bias towards vowels, as the vowel response distribution was 3 out of 9 choices, instead of 5 out of 7 as in the previous experiments.

The task consisted of 549 trials with no feedback. On 91.8% of all trials there were two targets. T2 followed T1 with 0, 2 or 7 distractors in-between (Lag 1, Lag 3 and Lag 8, respectively, and 39.3%, 26.2% and 26.2% of all trials, respectively). T1 was a solo target in 8.2% of all trials in which T1 could be a single formant or vowel; each of the 9 response alternatives was a solo target in 0.91% of all trials. The experiment lasted approximately 60 minutes.

Results and discussion

T1 accuracy was strongly affected by Lag, $F(2, 28) = 12.271$, $MSE = 0.003$, $p < 0.001$. Performance averaged 26% at Lag 1, compared to 32.3% at Lag 3, and 35.9% at Lag 8. When report order was ignored performance was 37.5% at Lag 1, 38.3% at Lag 3 and 42.3% at Lag 8. This is illustrated by the left panel of Figure 7.

The accuracy for (T2|T1) was affected by Lag, $F(2, 28) = 3.562$, $MSE = 0.011$, $p < 0.05$. Performance averaged 28.9% at Lag 1, compared to 33.9% at Lag 3, and 39.2% at Lag 8. A paired samples t-test showed a significant difference between Lag 1 and Lag 8 ($t(14) = -2.550$, $MSE = 0.040$, $p < 0.05$), again indicating an early attentional blink. When report order was ignored, performance was 45.4% at Lag 1, 37% at Lag 3 and 40.8% at Lag 8, as shown in the right panel of Figure 7.

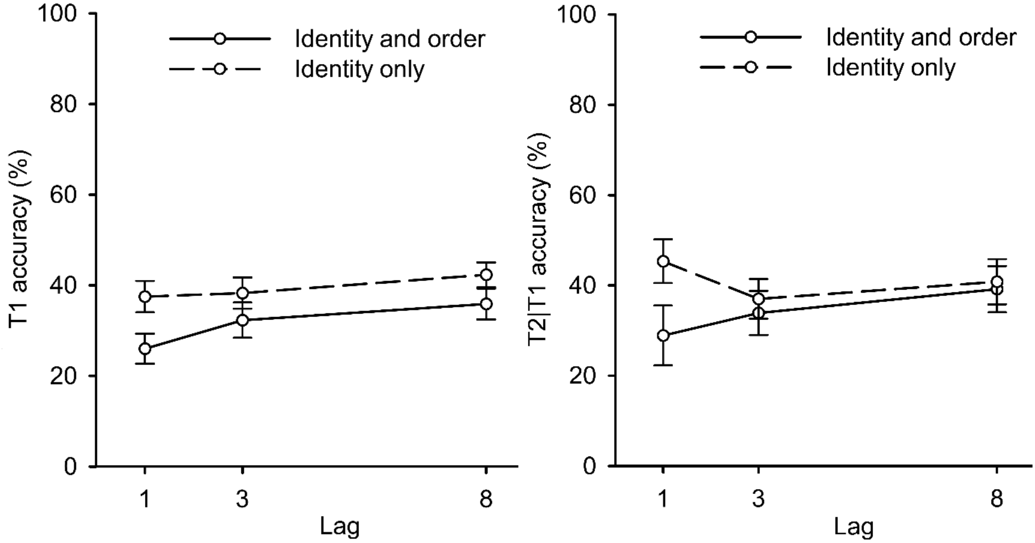


Figure 7. Experiment 3: The left panel shows T1 task performance for each lag. Error bars represent ± 1 standard error of the mean. The right panel shows (T2|T1) performance for each lag. Dashed lines represent identification accuracy if report order is ignored.

Importantly, the frequency of strict integrations was strongly affected by Lag, $F(1.1, 15.4) = 12.208$, $MSE = 0.082$, $p < 0.005$. Integrations averaged 35.3% at Lag 1, compared to 5% at Lag 3, and 0% at Lag 8. Order reversals were not affected by Lag, $F(2,28) = 0.483$, $MSE = 0.009$, $p = 0.622$. Reversals averaged 10.5% at Lag 1, compared to 9.2% at Lag 3, and 7% at Lag 8.

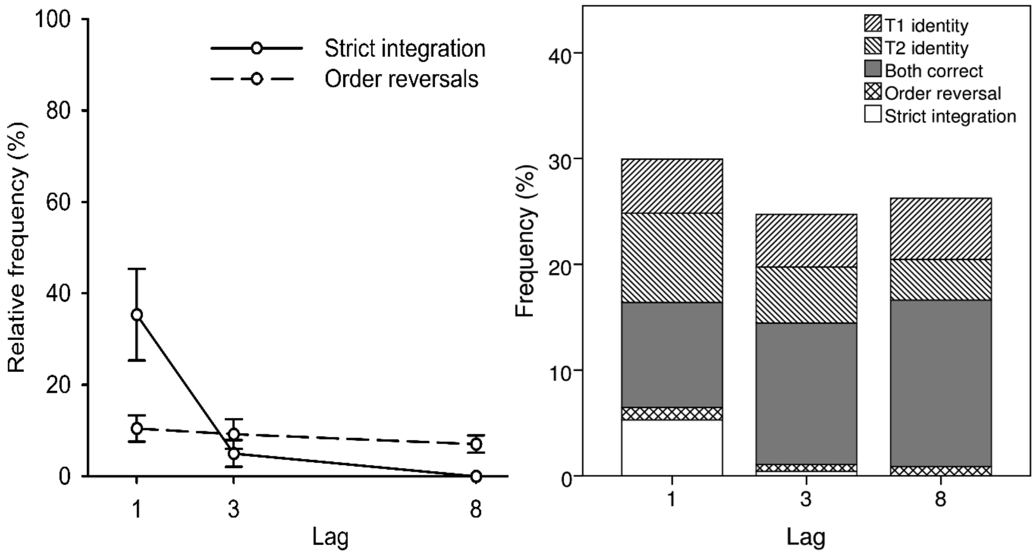


Figure 8. Experiment 3: The left panel shows the relative frequency of strict integrations and order reversals for each lag, as a percentage of the total number of responses in which both target identities were preserved. The right panel shows the distribution of responses for each lag, as a percentage of the total number of responses.

Figure 8 shows that despite the fact that Lag 1 was not completely dominated by strict integrations, as was the case with the previous two experiments, the number of strict integrations was still relatively high at Lag 1. Indeed, strict integrations were almost solely present at Lag 1. Integration of targets at longer intervals (Lag 3 and 8), and with multiple intervening distractors, was not necessarily predicted, and so the absence of integration reports at these longer lags was in line with expectations. The 'baseline' frequency of integration reports at these longer lags in the previous experiments is thus indeed likely to have resulted from a response bias towards vowels, which the present experiment removed. Importantly, however, at Lag 1, where integration is expected, the number of integrations remained substantial. The frequency of order reversals, on the other hand, still did not change across lags.

GENERAL DISCUSSION

The present study investigated whether two rapidly following auditory stimuli can be integrated and perceived as if they were presented simultaneously, resulting in a unitary integrated percept, similar to what is commonly observed in the visual domain. This was confirmed in three versions of an RSAP task. Participants indeed frequently only reported an integrated percept of a synthetic vowel at Lag 1, while such reports were rare at longer lags, consistent with the idea of temporal integration. The perception of a single synthetic vowel when two complex tones were presented non-simultaneously (at Lag 1) also shows that temporal integration is much more complex than simple energy summation, an interpretation previously given by some authors (Pedersen & Elberling, 1972; Pedersen & Salomon, 1977; Zwislocki, 1969). Conversely, the combined results of the present experiments also suggested that integration does not rely heavily on high-level (linguistic) knowledge either: Integration was as frequent with natural transitions as it was without.

The current findings are overall most compatible with more comprehensive accounts of temporal integration, such as discussed by Moore (2003), or Näätänen and Winkler (1999), except for the fact that they hypothesize that acoustic information is integrated and placed on temporal coordinates while the present data show that temporal information is often lost: When identity information of both targets was retained, participants frequently reported to hear only the integrated percept of a synthetic vowel, which was the correct assembly, but also the temporal merger of both target formants, instead of reporting both targets in the correct or incorrect order, despite the inherently high temporal resolution of the auditory system (Eddins & Green, 1995).

Interestingly, in cases when two targets were heard, order information did not seem to suffer from the temporal proximity of the targets. This contrasts with the findings obtained in visual tasks, which do show an increase in order errors at Lag 1, even if their frequency is relatively low overall (Akyürek et al., 2012). If anything, order errors were reduced at Lag 1, at least in Experiment 1 and 2, although this might also be a consequence of a reduced ability to separate the targets in the first place. At subsequent lags, when reports of integrated percepts decreased, there was a proportional increase in fully correct responses, while order reversals remained infrequent, but relatively constant across lags. When temporal integration does not occur, it thus seems the auditory system does keep close track of stimulus order.

Relationship to previous studies on tone perception

Findings from MMN studies may at first glance appear to contrast with the present results. However, although deviance detection in MMN studies seems to suggest that temporal order is retained within integrated percepts, this may not be a necessary assumption: Grouping pairs (or more) of stimuli together in one percept and dissociating it from other tones that occur after longer delays only requires that the integrated percept is perceived in time in reference to other percepts. It does not necessarily require that its constituent parts are also ordered correctly in time.

Some findings of Tervaniemi et al. (1994) provide some further support for this view. In their study, pairs of two different tones were presented in series, separated by silent gaps. During this continuous stream, when the second tone of a pair was omitted, an MMN was elicited. One might thus conclude that each tone pair was regarded as a unitary event and that the listener expected to perceive the first and second tone of the pair in order, as the definition says that integrated stimuli are placed on temporal coordinates (Näätänen & Winkler, 1999). Yet, this account seems inconsistent with the fact that no statistically significant MMN was elicited when Tervaniemi et al. (1994) reversed the first and second tone of the pair, instead of omitting the second tone. This suggests that a deviant, order-reversed tonal pair is not regarded as a deviance from the norm by the auditory system (per se). Although the absence of an MMN as such in this study may not be fully conclusive, the observed non-deviance of an order-reversed tone pair does suggest that order information within the perceptual event might have been missing.

Findings from another study conducted by Ciocca and Darwin (1999), focusing on pitch perception, might also support the idea that integrated auditory stimuli are not placed on temporal coordinates. In this study, non-simultaneous mistuned sound components presented temporally close to a target sound changed its perceived pitch. However, by themselves, these results can also be accounted for by assuming that pitch processes work from samples in short-term memory (as proposed in the multiple-looks model by Viemeister, 1996), creating a virtual pitch without losing the individuality and temporal order of the stored samples.

The current findings are, however, largely incompatible with the multiple-looks model, as it assumes that there is no true long-term integration of the kind that we report here (Viemeister, 1996). The multiple-looks model states that long-term temporal integration (of the magnitude presently obtained) can only be achieved by computations made on short-term samples (+/- 3 ms) in short-term memory. The present data clearly show that the information available for report (i.e., in short-term memory) does not seem to consist of individual samples; instead, these appear to have been lost or irreversibly overlaid by an integrated percept—otherwise there would be no reason for not reporting the individual targets in the present task. Recall that participants were in principle expecting to be able to report two targets, not just one, and that the integrated percept was in fact much rarer than the formants, thus making the latter unappealing as a response choice from a strategic perspective.

To accommodate the present results, the multiple-looks model could possibly be modified by allowing the computations that are assumed to apply to multiple samples in short-term memory to act as a kind of long-term temporal integration window (of a few hundred ms), which assembles the samples into a single acoustic percept at the expense of the individuality of the samples. However, this would seem to go against one of the principal tenets of the model, namely that integration across longer intervals does not take place (Viemeister, 1996).

Relationship to the continuity illusion and phonemic restoration

One might suspect that auditory temporal integration is related to the continuity illusion, which is the perception of a discontinuous, interrupted signal as being continuous when the gaps are filled by loud noise (Başkent, Eiler, & Edwards, 2009; Carlyon et al., 2002; Heinrich et al., 2008; Warren, Obusek, & Ackroff, 1972). The illusion is strongest when temporal and spectral components from the noise are matched to those from the sound signal (Bregman, 1994; Riecke, Opstal, & Formisano, 2008; Warren, 1999). During the continuity illusion, separate (interrupted) stimuli are perceived as one coherent signal, which is similar to temporal integration, as the current research showed that people can perceive two stimuli (separated by a temporal gap) as a single integrated entity. Furthermore, in a previous study, MMN latency data indicated that the processes underlying the continuity illusion are active within a period of 200 ms after the onset of the noise-filled gap, an interval that is comparable to that of temporal integration (Micheyl et al., 2003).

There is, however, no evidence that during the continuity illusion the continuously perceived entity is being integrated into a single, overlaid entity. To wit, tone sweeps gliding upward or downward in frequency can be perceived as continuous when they are interrupted and their silent gaps filled with noise (Ciocca & Bregman, 1987). In other words, people perceive the tone sweep to continue during the noise with a similar upward or downward trend as before the interruption occurred. Would there be a true integration, then a compound of tones with different frequencies might be a more likely percept instead. Furthermore, the continuity illusion for steady-state tones can still occur when the intervening noise is up to 2000 ms long (Riecke et al., 2008). Such a duration lies outside the scope of the temporal window of integration. Lastly, as these examples also show, for the continuity illusion to work, a filler stimulus is needed, such as noise, to bridge the silent gap. Temporal integration requires no such masker, which, in fact, might even impair integration.

Because of its more linguistic nature, a special case of the continuity illusion may be particularly relevant to temporal integration as presently tested: Phonemic restoration, which is the ability to perceptually restore and enhance intelligibility of interrupted, degraded speech (Başkent et al., 2009; Başkent, 2012; Warren & Sherman, 1974). Phonemic restoration is commonly observed with interrupted speech that has comparable speech and silent/noise intervals to that of temporal integration, but phonemic restoration is clearly more complicated, as it is an interaction between top-down and bottom-up factors, including expectations, linguistic skills, situational and semantic context, Gestalt rules, as well as spectral and temporal cues from the speech (Bashford, Riener, & Warren, 1992; Başkent, 2012; Davis & Johnsrude, 2007; Samuel, 1981; Stenfelt & Rönnberg, 2009). Nonetheless, in the current task, in order to identify the correct vowel after both formants are integrated, some knowledge of the vowels from the response alternatives was applied, and a possible role of attentional selection (or top-down control) seems feasible also.

There is indeed prior evidence for common ground between temporal integration and restoration of degraded speech. Using a speech restoration task, Saberi and Perrott (1999) showed that speech intelligibility was almost perfect when speech segments of 50 ms were reversed in time, and only decreased when segments of 100 ms were reversed. Nonetheless, when repeatedly listened to the latter condition, participants reported that the words gradually became clearer and easier to understand, and they eventually reported actually hearing the words. Apparently, while these segments were reversed in nature (i.e., temporally distorted), there was still enough information for the auditory system to reconstruct meaningful objects. Temporal coordinates might thus not be fully fixed, and may be re-ordered or re-interpreted if needed.

Although the results of Saberi and Perrott (1999) are intriguing, it seems likely that perceptual organization works differently for meaningful speech units, especially for context-rich sentences (Clarke, Gaudrain, Chatterjee, & Başkent, 2013), than with simpler auditory stimuli. Considering that some practice was needed in the study of Saberi and Perrott (1999), it seems likely that the perceptual reconstruction involved both a re-utilization of other speech cues that were not distorted, as well as top-down processes, such as expectancies, linguistic skills, and vocabulary, to correctly interpret the distorted speech signal (Bashford et al., 1992; Başkent, 2012; Davis & Johnsrude, 2007; Samuel, 1981). In other words, perhaps their results would have been different if, instead of highly redundant speech, simpler speech materials were used, such as vowels, syllables, or words without context. The present results address such doubts to an extent: Temporal integration as measured in the present task seems to confirm that temporal coordinates may not always play an important role in the perception of brief events.

Conclusion

When successive, broadly compatible tones are perceived across an interval of up to 200 ms, temporal integration of these stimuli frequently may give rise to a unified percept that consists of featural properties of the individual tones, but which (strongly) diminishes their individuality and temporal properties. Thus, temporal integration in the auditory domain is similar to that observed in vision, supporting the view that temporal integration may be a general, amodal perceptual processing function in the human brain.

CHAPTER 3

Visual and auditory temporal integration in healthy younger and older adults

J. D. Saija, D. Başkent , T. C. Andringa and E. G. Akyürek

Published in Psychological research (2017). doi: 10.1007/s00426-017-0912-4.

ABSTRACT

As people age, they tend to integrate successive visual stimuli over longer intervals than younger adults. It may be expected that temporal integration is affected similarly in other modalities, possibly due to general, age-related cognitive slowing of the brain. However, previous literature does not provide convincing evidence that this is the case in audition. One hypothesis is that the primacy of time in audition attenuates the degree to which temporal integration in that modality extends over time as a function of age. We sought to settle this issue by comparing visual and auditory temporal integration in younger and older adults directly, achieved by minimizing task differences between modalities. Participants were presented with a visual or an auditory rapid serial presentation task, at 40-100 ms/item. In both tasks, two subsequent targets were to be identified. Critically, these could be perceptually integrated and reported by the participants as such, providing a direct measure of temporal integration. In both tasks, older participants integrated more than younger adults, especially when stimuli were presented across longer time intervals. This difference was more pronounced in vision, and only marginally significant in audition. We conclude that temporal integration increases with age in both modalities, but that this change might be slightly less pronounced in audition.

INTRODUCTION

Stimuli that rapidly succeed one after another can be perceived as a single composite stimulus and/or event. When watching a movie, for example, rapid, successive still images are perceived as fluent motion. This is due to a perceptual process named temporal integration, which combines stimuli within an interval up to about 200 ms into an aggregated representation (Hogben and Di Lollo, 1974; Di Lollo, 1980). The duration of the interval varies from person to person, however, and factors that affect cognitive functioning can play a role therein. A person's age, then, can be an important factor, since aging results in an overall decline or slowing down of the cognitive system (Salthouse, 1996). Yet, how aging affects temporal integration specifically is not yet fully known.

In vision, several studies on temporal integration of visual forms have shown that older adults visually integrate across longer time intervals. For instance, Di Lollo, Arnett, and Kruk (1982) presented participants with two 5x5 dot matrices, presented simultaneously side by side, but with the successively plotted dots presented for just 1.5 µsec. Participants were asked which of the two matrices contained a missing dot (Di Lollo, Arnett and Kruk, 1982). To find the missing dot, it is necessary to temporally integrate all dots as if they were presented simultaneously, because consolidating, let alone mentally comparing, 25 positions in such a short time would be impossible. The authors varied the total plotting interval by adjusting the interstimulus interval (ISI) between dots, and found that the younger group needed a shorter plotting interval (60.5 ms) to obtain the same level of 75% task performance as the older group (85 ms). This suggests that the older group temporally integrated the individual sequential dots over a longer interval than the younger group, indicating a longer temporal integration window.

Converging evidence has also been obtained with different tasks, such as color integration (fusion). Kline and colleagues (1982) briefly presented participants with a green circle followed by a red circle, both presented in the same location. Perceptually overlaying both circles would result in perceiving a yellow circle. By varying the ISI between the two circles, the authors could measure within what time window participants would temporally integrate the green and red circles and resultantly perceive a yellow circle. The authors found that the older group reported perceiving more color integrations up until the longest ISI, which amounted to a total stimulus duration of 90 ms. The younger group, in contrast, only reported seeing color integrations up to a total stimulus duration of 70 ms. Similarly, in a word recognition task, Kline and Orme-Rogers (1978) measured performance for three-letter words consisting of horizontal and vertical lines, by displaying two halves of random lines of each individual word sequentially. Recognizing the words becomes possible when a participant temporally integrates both halves in a single perceptual representation, which becomes easier when the ISI is small. Across a total stimulus duration range of 100-200 ms, the authors found that the older participants had higher word recognition scores with longer ISIs than the younger participants, which can be explained by a longer temporal integration window for the older group.

As alluded to, one explanation to why aging leads to increased visual temporal integration can be age-related cognitive slowing. According to the processing-speed theory, cognitive slowing would lead to carrying out fewer cognitive operations within a certain timeframe (Salthouse, 1996; Madden and Allen, 2015). When time is limited or processing time is externally constrained, later cognitive operations are then left with less processing time as earlier operations are taking longer to finish. Additionally, due to cognitive slowing, memory traces of the results of earlier operations may decay before they can be used for later operations, which illustrates that cognitive slowing causes substantial 'collateral damage' apparent as noticeable impairments in daily life activities.

Given the fairly consistent results in the visual domain, one might expect that the auditory modality should be similarly affected. The supposed global nature of cognitive slowing is also compatible with that idea. To wit, measures reflecting other temporal aspects of vision and audition indeed change similarly with age: For both vision and audition, older adults have higher gap detection thresholds (Humes et al., 2009) and are more susceptible to backward masking (Di Lollo, Arnett and Kruk, 1982; Gehr and Sommers, 1999). However, to our knowledge, there are no studies that have provided direct evidence that the auditory temporal integration window is longer for older adults. In fact, there is indirect evidence pointing to the contrary. An electroencephalographic study on the mismatch negativity (MMN; elicited by a violation in a to-be-expected order or identity of repetitive stimuli; Näätänen, Kujala, and Winkler, 2011) showed that the duration of the auditory temporal integration window does not differ between younger and older adults (Horváth et al., 2007). Using two kinds of oddball experiments (double deviant and stimulus omission), the authors showed that the temporal integration window of their younger participants was around 250 ms, and the window of the older participants was around 200 to 250 ms.

The lack of evidence for prolonged auditory temporal integration leaves the possibility that aging might be affecting temporal integration differently for each sensory modality. The degree to which integration changes with aging might depend on the relative importance of time in a given sensory modality. In the visual modality, for instance, space is more dominant than time, and it is conceivable that the functionally weakest neurons (i.e., those dealing with temporal aspects) are the first to atrophy when people age. Analogous effects are seen in the body when age-related muscle atrophy is observed (Abate et al., 2007); the so-called “use it or lose it” principle (Schooler, 2007). In perception, the principal dimension of vision is space, but the principal dimension of audition is time (Kubovy, 1988; O’Callaghan, 2008). For example, the borders of visual objects are inherently indicated by coordinates in space, while those of auditory objects are defined in time. Also, it is easier to imagine an object that is independent of time in the visual domain (e.g., a still image) than in the auditory domain. In line with these conjectures, Humes and colleagues (2009) showed that auditory gap detection thresholds are lower than the visual ones, and that age differences appear to be larger for visual than for auditory stimuli.

Apart from a general effect of time, temporal integration might also be spared more specifically, because temporal integration is required on a daily basis to process and understand speech (Poeppel, 2003; Wallace and Blumstein; 2009): Especially to analyze vowels, higher-level processes map auditory information within 200 ms onto linguistic representations in the form of a phonetic category decision. Also, even though research showed that older adults have more difficulties with understanding speeded speech (Wingfield, 1996; Gordon-Salant and Fitzgibbons, 2001), Schneider et al. (2005) showed that auditory decline and speed-induced stimulus degradation, but not cognitive slowing, may be responsible for lower intelligibility. Thus, it remains conceivable that age-related decline in temporal processing and integration might be lessened in the auditory domain.

Current research

Taken together, there is substantial evidence indicating that aging increases visual temporal integration, but for the auditory domain the picture is less clear. Two possibilities exist: First, temporal integration may occur over longer intervals for the older population regardless of the specific sensory modality, which would seem compatible with the notion of general cognitive slowing. Second, differential aging effects on temporal integration in each modality may occur. Such a finding would suggest that the "use it or lose it" principle may apply, meaning that the visual modality could be affected by aging more than in the auditory modality, because the time dimension is less important in vision compared to the space dimension.

The main purpose of the present study was thus to investigate whether aging similarly affects temporal integration in both the visual and auditory domain. Clear evidence from a cross-modality comparison can only be provided with a task that provides a direct measure of temporal integration in each modality equally. In the present study, the visual and auditory tasks were made as similar as possible, by using the rapid serial visual presentation (RSVP; Akyürek et al., 2012) task and its auditory equivalent, rapid serial auditory presentation (RSAP; Saija et al., 2014a). For each task we tested multiple stimulus durations (40, 70 and 100 ms). If aging affects temporal integration, then this should be reflected in older adults reporting more temporal integration for longer stimulus durations when two targets succeed each other directly (i.e., at Lag 1), in particular. More specifically, the number of temporal integration reports for older adults should decrease at a lower rate with longer stimulus durations compared to younger adults. This should then be reflected in a significant interaction effect of age and stimulus duration.

EXPERIMENT 1A: VISUAL TEMPORAL INTEGRATION

Methods

Participants

Participants were naive to the purpose of the experiment. Since the experiment relied on visual stimuli, all participants were confirmed to have normal or near-normal vision according to the Ranges of Vision Loss established by the International Council of Ophthalmology in 2002. The participants' visual acuity was measured (with lenses or glasses if required) using the Landolt C test. The mean visual acuity for the young group was LogMAR -0.16 and for the older group LogMAR -0.02. Fig. 1 shows the visual acuity as a function of age. Further, mental flexibility and normal cognitive functioning were confirmed with the Trail Making Test Part A and B (Chanmugam, Triplett and Kelen, 2013). Three older adults were excluded from participation because one was suffering from macula pucker, one was unable to perform the task, and one had a stroke in the past. After exclusion, nineteen young students of the University of Groningen (6 male, 13 female) with a mean age of 20 years (ranging from 17 to 23 years) and nineteen older adults (16 male, 3 female) with a mean age of 70 years (from 65 to 81 years) participated in the study. Younger participants received course credit or monetary compensation, while older participants only received monetary compensation. Informed consent was obtained in writing before participation, and the study was approved by the Ethical Committee of the Department of Psychology at the University of Groningen.

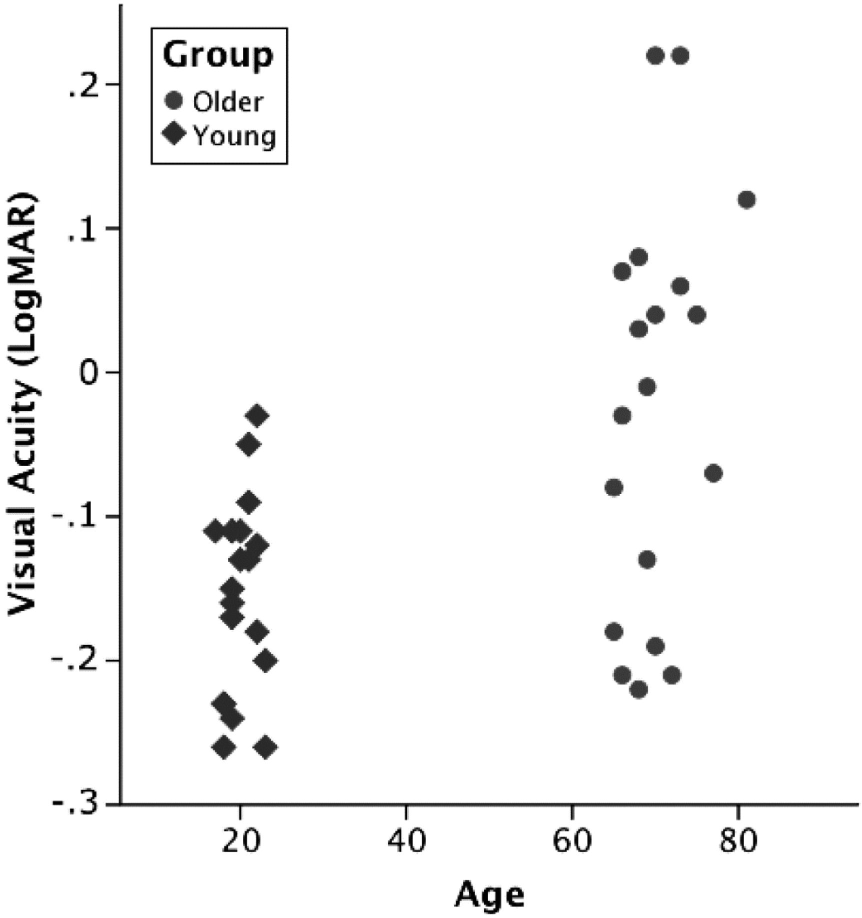


Figure 1 - Experiment 1A: This graph shows the visual acuity in LogMAR for young and older participants by age.

Apparatus and stimuli

The experiment was implemented with E-Prime Professional 2.0.8.90 (Psychology Software Tools, Pittsburgh, PA) running on a desktop computer with Microsoft Windows XP. The visual stimuli were presented on a 19-inch CRT screen, which refreshed at 100 Hz with a resolution of 1024x768 pixels in 16-bit color, and which was placed at a viewing distance of approximately 60 cm. The participants' responses were collected via a keyboard.

The target stimuli consisted of the symbols **/ \ o** and their combinations, as shown in Fig. 2. They were at most 49 pixels in height and 33 pixels in width (approximately 1.6° and 1.1° degrees of visual angle, respectively) and were displayed in red (RGB 255, 0, 0; 91 cd/m^2). The targets were chosen such that their features did not overlap with each other (e.g., the **/** was never presented with the **X**). The distractor stimuli were drawn without replacement from the modern Latin alphabet (excluding I, J, K, L, O and X to avoid confusion with the target symbols). The distractor stimuli, as well as the fixation cross, were all printed in bold 52 pt. Courier New font and colored in black (RGB 0, 0, 0; 2 cd/m^2). The targets and distractors were about equal in size. The background color was always light gray (RGB 192, 192, 192; 265 cd/m^2).

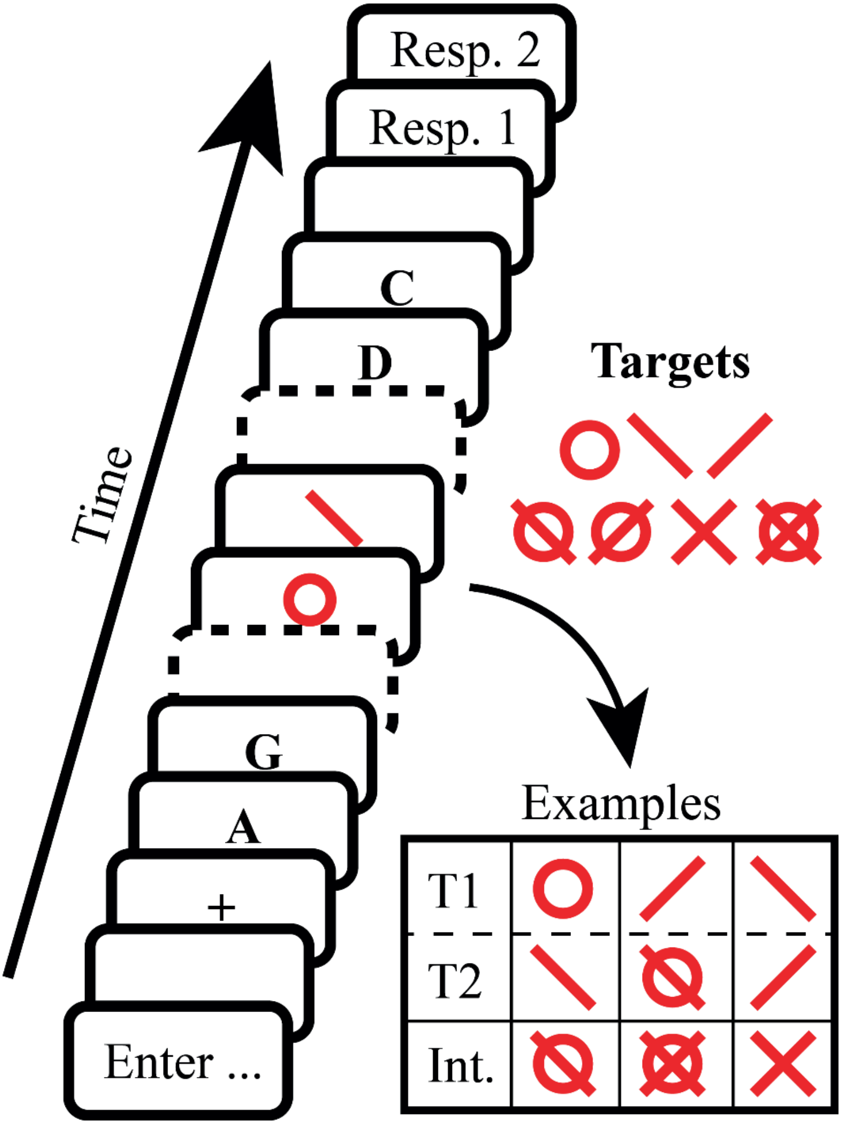


Figure 2 - Experiment 1A: Example of a typical trial to illustrate the procedure and visual stimuli. The empty boxes with solid lines represent blank periods of 100 ms. The empty boxes with dashed lines represent the succession of multiple distractor stimuli (i.e., black letters). The target stimuli were always presented in red. For each trial, all stimuli were of equal duration and were presented for 40, 70 or 100 ms. Each stimulus was separated by an ISI of 10 ms.

Procedure

The experiment consisted of a short block of practice trials and continued with 496 experimental trials with an optional break halfway, lasting for approximately 60 to 90 minutes. At 100 ms after a trial was initiated by a participant, the fixation cross was displayed for 200 ms. Then 19 stimuli succeeded each other, all of which were on screen for 40, 70 or 100 ms and followed by a 10 ms blank screen each (50, 80 and 110 ms SOA respectively; 1/3 of trials each). On 94.4% of the trials, two of these stimuli were targets (T1 and T2), while the others were distractors. T1 appeared as either the 5th or the 7th item in the stream and T2 followed T1 with either 0, 2, or 7 distractors in-between, referred to as Lag 1, 3, or 8 (31.5% of trials each). On 5.6% of the trials T1 was a solo target.

The participants were told that each trial could contain one or two targets, and they were asked to identify each of them. After each stream, a 100 ms blank screen was presented, after which the participants were asked to enter the identity of T1 and then that of T2 on the numerical keypad. Each target response alternative was labeled on the numerical keypad. If a target was not spotted, then an empty response could be given by pressing the Enter key. However, guessing was encouraged when a participant was unsure about the identity of a target. Fig. 2 shows an example of a trial that illustrates the procedure.

Analyses

Of main interest were the reports of integrated percepts (i.e., reporting the integrated percept of the combined features of T1 and T2) that were reported as a single response (i.e., no second response was entered). These responses were regarded as strict integrations, and indicated that the observer only perceived a single target, which constituted of the integrated combination of T1 and T2. Secondly, task performance was analyzed, which reflects correct response accuracy of the target identities and their temporal order. Analyses were performed on the number of trials in which T1 was correctly reported, and in which T2 was correctly reported given that T1 was correct as well (T2|T1). T1 was also considered correctly reported when the integrated percept of T1 and T2 was reported (as was T2|T1).

The data were in the form of count data, and because the variance of the data for each analysis was larger than the data's mean, all data best fitted the negative binomial distribution. Therefore the data were analyzed using Generalized Estimating Equations using a negative binomial distribution with log link. For each analysis separately, the overdispersion parameter (α) was estimated and the working correlation matrix (WCM) was chosen based on the best goodness of fit (i.e., lowest quasi likelihood under the independence model criterion (QIC); Pan, 2001). Each analysis included the two within-subject variables stimulus duration (40, 70 and 100 ms) and T1-T2 Lag (1, 3, and 8), as well as the between-subject variable age group (young and older participants). Strict integrations were expected to happen mostly at Lag 1 due to the short distance between targets and the lack of distractors in-between, and therefore additional analyses were performed on the data of Lag 1 only, whereby T1-T2 Lag was removed as a variable. For each test, a significance level of .05 was used.

The strict integration reports were represented as relative frequencies, that is, relative to all trials in which both target identities are retained regardless of their positions (i.e., strict integrations, order reversals and both correct responses). For reference, the Appendix contains figures with the absolute integration rates for all experiments reported here. To account for this relativity, the offset for each combination of subject and condition was included in these analyses and was calculated as the natural log of the exposure (i.e., of the number of trials that include strict integrations, order reversals and both correct responses, per subject and condition). For the (T2|T1) accuracy, the offset for each combination of subject and condition was calculated as the natural log of the exposure of the number of trials in which T1 was correct. For T1 accuracy there was no relativity, so for each subject and condition all trials could be included. Therefore the T1 offset for all conditions and subjects was set to the natural log of the total number of trials per condition and subject ($\ln(52) \approx 3,95$).

The estimated marginal means of the analyses of relative frequencies of strict integration reports were plotted in bar graphs. The estimated marginal means of the T1 and (T2|T1) accuracies were also plotted in bar graphs, together with the accuracies when report order is ignored (e.g., when T1's identity is correct regardless of T1's position, namely including T1 reported as T2, order reversals and strict integrations).

Results

A full factorial analysis (WCM = autoregressive, α = 15.322) was performed on the relative frequencies of strict integration (i.e., relative to strict integrations, order reversals and both correct responses), which are shown in Fig. 3. The frequency of strict integrations was significantly affected by lag, $\chi^2(2, N = 342) = 64.7$, $p < 0.001$, by stimulus duration, $\chi^2(2, N = 342) = 95.5$, $p < 0.001$, and by their interaction lag*duration, $\chi^2(4, N = 342) = 55.5$, $p < 0.001$. Fig. 3 shows that reports of strict integrations are most prominent at Lag 1 and become less frequent with longer lags and longer stimulus durations. Strict integrations were also affected by group, $\chi^2(1, N = 342) = 19.6$, $p < 0.001$, as well as by the interactions of group*lag, $\chi^2(2, N = 342) = 8.8$, $p < 0.015$, and group*duration, $\chi^2(2, N = 342) = 21.8$, $p < 0.001$.

An additional analysis for Lag 1 only (WCM = unstructured, α = 3.029), showed that stimulus duration was a significant factor, $\chi^2(2, N = 114) = 68.9$, $p < 0.001$, which indicates that shorter stimulus durations resulted in more reports of strict integrations. Even more importantly, older adults were more influenced by stimulus duration than young adults, revealed by an interaction effect of group*duration, $\chi^2(2, N = 114) = 26.7$, $p < 0.001$, meaning that older adults integrated more often than young adults at longer stimulus durations. Also, older adults reported more strict integrations at Lag 1 over all three durations, $\chi^2(1, N = 114) = 17.6$, $p < 0.001$. These effects can be seen in more detail in Fig. 3.

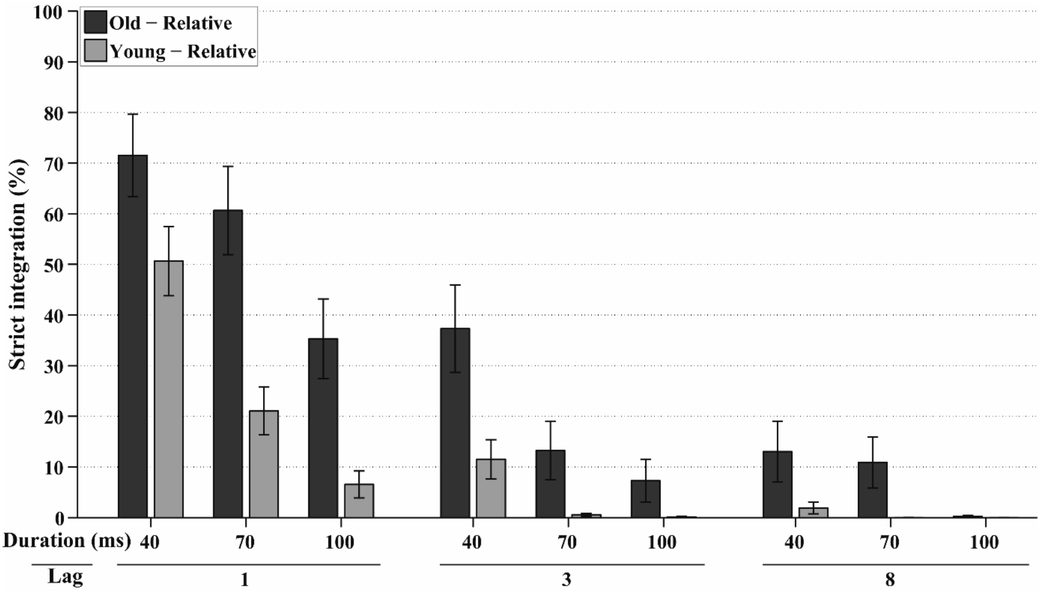


Figure 3 - Experiment 1A: This graph shows the estimated marginal means of the analyses of relative frequency of strict integrations for all combinations of stimulus duration, lag, and age group, as a percentage of the total number of trials in which both target identities were preserved. Error bars represent ± 1 standard error of the mean.

Another factorial analysis (WCM = unstructured, α = 2.015) was performed on the frequency of trials where T1 was correct. The average accuracy of T1 per group and for each lag and stimulus duration can be seen in Fig. 4, together with the average accuracy when report order is ignored (i.e., relaxed criterion). T1 accuracy was significantly affected by lag, $\chi^2(2, N = 342) = 213.3$, $p < 0.001$, and stimulus duration, $\chi^2(2, N = 342) = 137.8$, $p < 0.001$, as well as by their interaction lag*duration, $\chi^2(4, N = 342) = 43.3$, $p < 0.001$. Fig. 4 reveals that T1 accuracy was higher for each stimulus duration when lags were longer, as well as for each lag when the stimulus durations were longer. The accuracy of T1 also differed per age group, $\chi^2(1, N = 342) = 21.4$, $p < 0.001$, indicating that the younger group overall had higher performance. Also, group*lag was significant, $\chi^2(2, N = 342) = 8.6$, $p < 0.015$, as well as group*lag*duration, $\chi^2(4, N = 342) = 9.9$, $p < 0.045$.

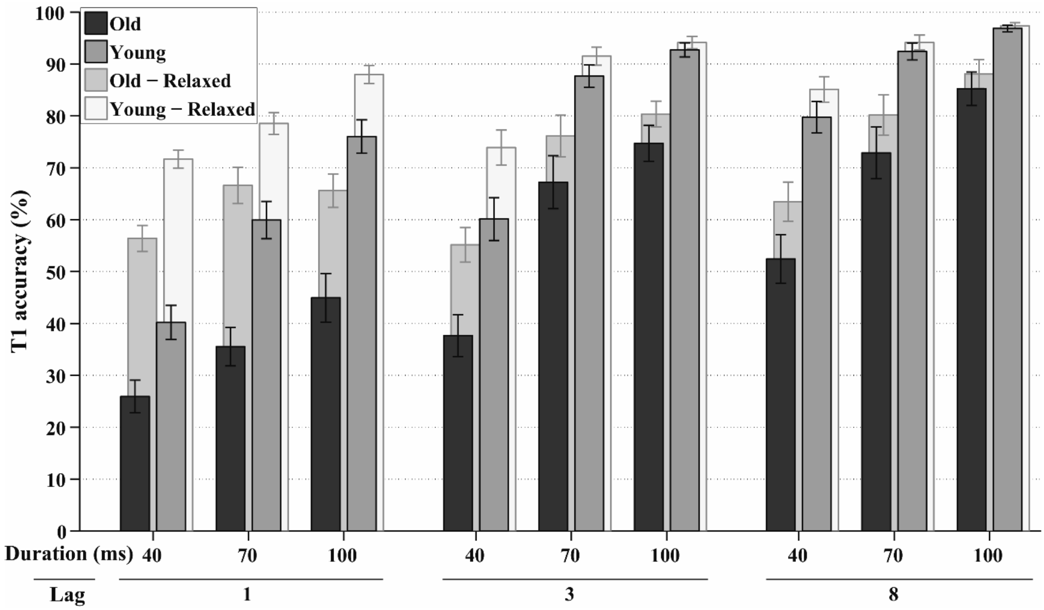


Figure 4 - Experiment 1A: The solid bars at the front show the estimated marginal means of the analyses on T1 task performance in percent correct, plotted for all combinations of stimulus duration and lag, for both age groups. The transparent bars at the back show the same analyses if report order is ignored (i.e., relaxed accuracy criterion). Error bars represent ± 1 standard error of the mean.

A final full factorial analysis (WCM = independent, α = 2.667) was performed on the number of trials where T2 was correct given that T1 was correct as well (T2|T1). Fig. 5 shows the average accuracy of T2|T1 per group and for each lag and stimulus duration, as well as the average accuracy when report order is ignored. T2|T1 accuracy was significantly affected by lag, $\chi^2(2, N = 342) = 108.5$, $p < 0.001$, and stimulus duration, $\chi^2(2, N = 342) = 138.8$, $p < 0.001$, as well as by their interaction lag*duration, $\chi^2(4, N = 342) = 36.6$, $p < 0.001$. Fig. 5 reveals that T2|T1 accuracy was higher for each longer lag or longer stimulus duration. The accuracy of T2|T1 also differed per age group, $\chi^2(1, N = 342) = 22.1$, $p < 0.001$, indicating that the younger group overall performed better. Also group*lag was significant, $\chi^2(2, N = 342) = 9.7$, p < 0.01, as well as group*duration, $\chi^2(2, N = 342) = 9$, $p < 0.015$, and group*lag*duration, $\chi^2(4, N = 342) = 14.7$, $p < 0.01$.

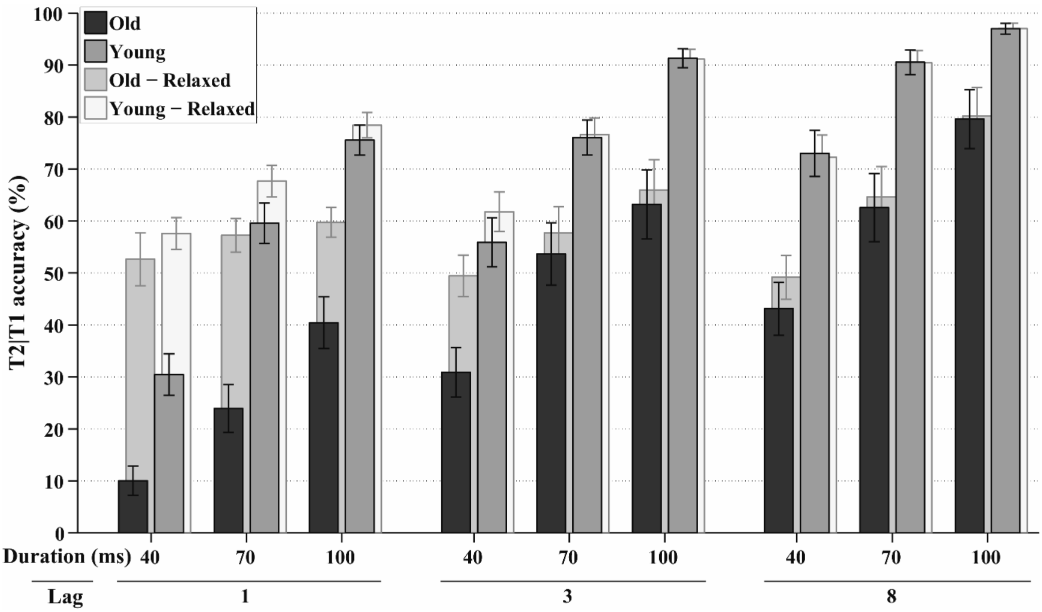


Figure 5 - Experiment 1A: The solid bars at the front show the estimated marginal means of the analyses on T2|T1 task performance in percent correct, plotted for all combinations of stimulus duration and lag, for both age groups. The transparent bars at the back show the same analyses if report order is ignored. Error bars represent ± 1 standard error of the mean.

Summarizing, older adults showed more integration than younger adults for visual stimuli, particularly for the longer stimulus durations tested. Elevated integration frequency was even observed at Lag 3, when 40 ms stimulus duration was used, for the older adults. For them, the speed of presentation seemed to overcome the inhibitory effects on integration of the intervening distractors. The younger group rarely integrated at Lag 3, even at the fastest presentation speeds. General task performance of the older adults, as measured by both T1 and T2|T1 accuracy, was also lower than that of the younger adults. Overall the results were thus in line with expectations.

EXPERIMENT 1B: THE EFFECT OF RETINAL ILLUMINANCE ON VISUAL TEMPORAL INTEGRATION

To be able to interpret the results of Experiment 1A unambiguously, it is necessary to exclude the possibility that the observed age-related differences could be due to purely sensory factors, such as increasing opacity of the lens with age. Specifically, it is conceivable that older people integrate more because their retinal illuminance is reduced (Coltheart, 1980; Di Lollo, Hogben and Dixon, 1994). Older people have on average a reduction of around a 0.5 log unit of retinal illuminance compared to that of younger people (Weale, 1963). To investigate whether the older adults in Experiment 1A perceived more integrated stimuli because of an inverse intensity effect (i.e., more integration with dimmer stimuli), a new group of younger adults was tested with 34% screen brightness instead of 100% in Experiment 1B, which simulates an approximate 0.5 log unit reduction in retinal illuminance. The experiment was otherwise identical to Experiment 1A (young group only).

Methods

Participants

Twenty-three young students of the University of Groningen (20 male, 3 female) with a mean age of 20 years (from 17 to 34 years) participated. All participants had normal or near-normal vision: The mean visual acuity for this new group of young adults was LogMAR -0.14. Fig. 6 shows visual acuity as a function of age. All participants received course credit for their participation.

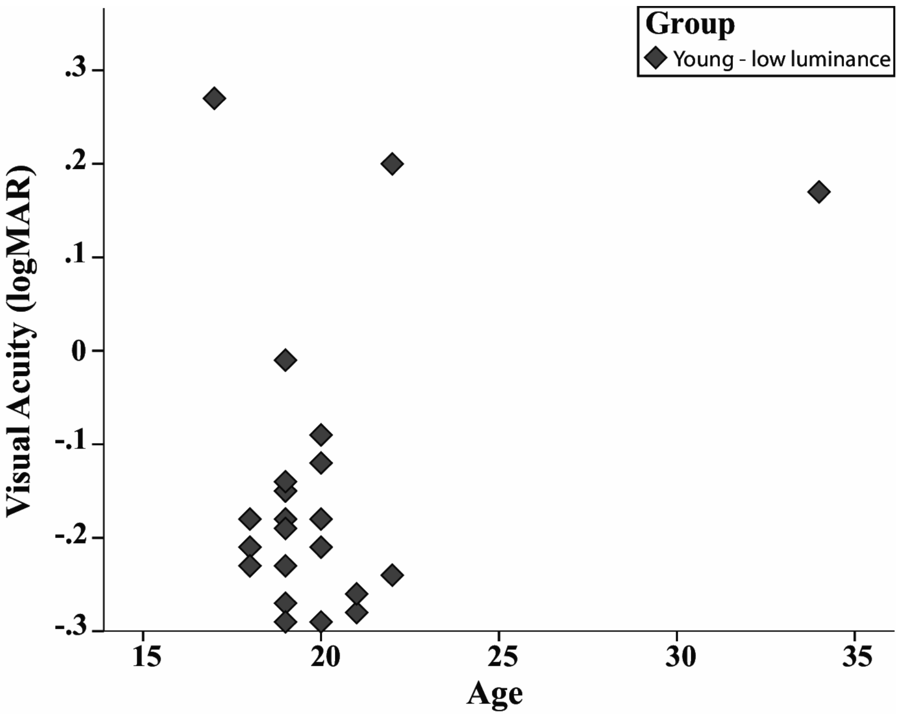


Figure 6 - Experiment 1B: This graph shows the visual acuity in LogMAR for the participants as a function of age.

Apparatus and stimuli

The only difference with Experiment 1A was that the brightness of the screen was set to 34% instead of 100% (in hardware), simulating reduced retinal illuminance as might be experienced by older observers. The red target stimuli were now displayed at 39 cd/m^2, and the light gray background at 109 cd/m^2.

Analyses

The analyses were focused on relative frequencies of strict integration reports. Firstly, we tested whether the reduced brightness in Experiment 1B resulted in more strict integration reports than in Experiment 1A, therefore the main analysis included the between-subject variable group (comparing the young participants from Experiment 1A with those from Experiment 1B), and the two within-subject variables stimulus duration (40, 70 and 100 ms) and T1-T2 Lag (1, 3, and 8). Secondly, a detailed analysis was performed on Lag 1 with the within-subject variable stimulus duration, for both the young participants of Experiment 1A and 1B.

Results

A full factorial analysis (WCM = autoregressive, α = 24.360) was performed on the relative frequencies of strict integration, which are shown in Fig. 7. The frequency of strict integrations was significantly affected by lag, $\chi^2(2, N = 378) = 100.387$, $p < 0.001$, by stimulus duration, $\chi^2(2, N = 378) = 30.09$, $p < 0.001$, and by their interaction lag*duratin, $\chi^2(3, N = 378) = 19.41$, $p < 0.001$. Fig. 7 shows that reports of strict integrations were most prominent at Lag 1 and became less frequent with longer lags and longer stimulus durations, as observed previously. Strict integrations were also affected by the interaction of group*duration, $\chi^2(2, N = 378) = 7.82$, $p < 0.025$, and the interaction of group*lag*duration, $\chi^2(2, N = 378) = 7.05$, $p < 0.035$, reflecting that low luminance seemed to decrease integration frequency in some conditions only, particularly at Lag 1, and at 40 ms duration.

An additional analysis for Lag 1 only (WCM = unstructured, α = 31.216), showed that only stimulus duration was a significant factor here, $\chi^2(2, N = 126) = 101.29$, $p < 0.001$. The lack of a significant group factor indicates that luminance did not have a significant effect on strict integration reports.

Even though Experiment 1B could not perfectly match the retinal illuminance of older observers (e.g., due to constant room lighting), the reduction in screen luminance was substantial enough that a sensory-driven rise in integration should have been revealed. However, the findings did not at all support the idea that reduced retinal illuminance might have fostered integration in the current task. As Fig. 7 shows, there was actually a trend in the opposite direction: Reduced brightness resulted in the perception of fewer integrated stimuli. Therefore, we can conclude that older people do not temporally integrate more because they perceive less brightness. The nature of the present task, in which dark stimuli appear on a light background (i.e., with inverse contrast), might have played a mediating role therein. Additionally, it is conceivable that a reduced ability to perceive darker targets may actually have limited the opportunity to integrate, as integration requires at least the perception of the stimulus features.

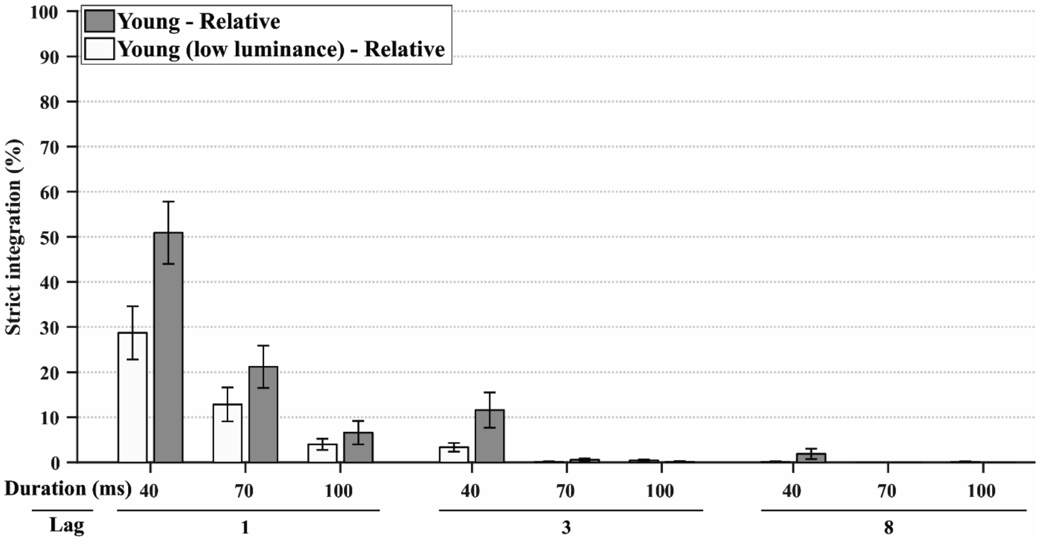


Figure 7 - Experiment 1B: This graph shows the estimated marginal means of the analyses of relative frequency of strict integrations for all combinations of stimulus duration, and lag, as a percentage of the total number of trials in which both target identities were preserved. The data from the young group of Experiment 1A (full luminance) are re-plotted next to the low luminance group for reference. Error bars represent ± 1 standard error of the mean.

EXPERIMENT 2: AUDITORY TEMPORAL INTEGRATION

The auditory Experiment 2 was carried out after Experiment 1, its visual counterpart, produced the expected pattern of results. It was similar to the RSAP experiment described in Saija et al. (2014a) but with two additional stimulus durations (40 and 70 ms). Similarly to the RSVP experiments, during the RSAP experiment a participant was presented with a stream of auditory instead of visual targets and distractors. The participant then had to report which targets were heard. The two auditory targets consisted of complex tones, which could be integrated pairwise into 2-formant synthetic vowels, analogous to the visual target combinations that were enabled in the RSVP experiments. During a pilot study with older participants it became clear that they were unable to discriminate between the original target stimuli and remember them, maybe as a result of age-related changes in temporal fine structure processing (Füllgrabe, 2013), age-related short-term memory deficits (Chen and Naveh-Benjamin, 2012), or some loss of auditory acuity (even if within the range of normal hearing; Martini, 1996). Therefore, the stimuli were modified in such a way that the older participants could discriminate the target stimuli more easily (as detailed below).

Methods

Participants

Participants were naive to the purpose of the experiment. Since the experiment relied on auditory stimuli, all participants were selected to have normal or near-normal hearing. They reported to have normal hearing, and their audiometric thresholds were tested using the definition of normal hearing from Martini (1996), namely that the 4-tone pure average across .5, 1, 2 and 4 kHz should be 20 dB HL or lower. Fig. 8 shows the audiometric thresholds for each individual for both age groups. Also, all participants were required to take the Trail Making Test Part A and B to test for mental flexibility and normal cognitive functioning (Chanmugam, Triplett and Kelen, 2013). An additional requirement was to be a fluent speaker of Dutch, as the stimuli were based on Dutch vowels. Two young and seven older participants were excluded from participation because they found the training too difficult. Also, eight older participants were excluded due to insufficient hearing, and two were excluded because they were unable to successfully finish the Trail Making Test Part B. After exclusion, twenty-two young students of the University of Groningen (11 male, 12 female) with a mean age of 20 (from 18 to 26) participated in the experiment for course credit. Also, twenty-two older adults (7 male, 16 female) with a mean age of 65 (from 60 to 71) participated for monetary compensation. Informed consent was obtained in writing before participation, and the study was again approved beforehand by the Ethical Committee of the Department of Psychology at the University of Groningen.

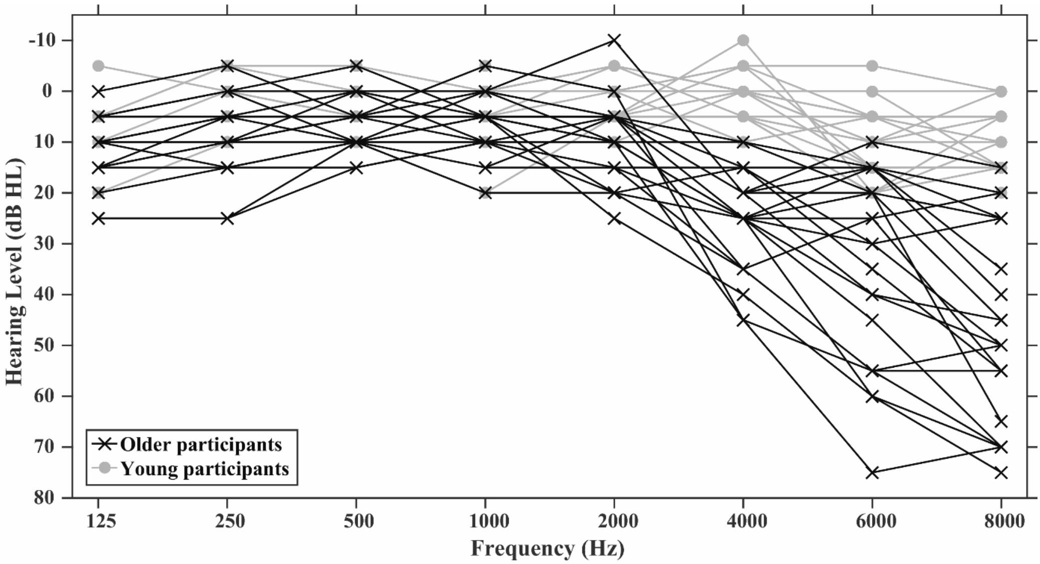


Figure 8 - Experiment 2: This graph shows the auditory acuity for the young and older participants in dB Hearing Level per frequency, plotted for the ear with the lowest hearing levels for each participant.

Apparatus and stimuli. The experiment was implemented in Matlab (8.5.0.197613; R2015a) using Psychtoolbox (3.0.12; Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) running on a Mac Pro with Mac OS X (10.10.4). Auditory stimuli were presented diotically through a Sennheiser HD 600 headphone, connected to an Echo Audiofire 4 external soundcard and a Lavry Engineering DA10 digital-to-analog converter. Responses were collected with a standard keyboard. Participants were tested in a sound-isolated booth.

The stimuli were created in Praat using a Klattgrid (Weenink, 2009), which is a speech synthesizer based on the Klatt synthesizer (Klatt, 1980; Klatt and Klatt, 1990). The Klattgrid program was used to create three Dutch vowels /a/, /i/ and /ø/ (Pols, Tromp and Plomp, 1973) with a pitch tier of 120 Hz, as well as the distractor tone, which was always the same and repeated during the experiment. Each vowel consisted of the first four formants (F1-F4; see Table 1). The use of four formants instead of two as in Saija et al. (2014a) ensured that the artificial vowels sounded more rich and more similar to natural vowels, making them easier to recognize and to discriminate between them. Each artificial vowel was divided in two parts, and each part was a possible target sound. One part contained F1 and F3, and was perceived as being lower in timbre than the distractor because most energy was at F1. The other part contained F2 and F4, and was perceived as being higher in timbre as most energy was at F2. F1 was lower in frequency than the distractor and F2 was higher (see Table 1). The bandwidth of F1 was set to 50 Hz, and the bandwidth of each subsequent formant was enlarged by 50 Hz compared to the previous formant. Part 1 was set at 65 dB SPL and each second part was set at a lower intensity (see Table 1) that would result in the best perception of the artificial vowel when both parts are combined. Additionally, a ramp of 5 ms was placed at each on- and offset to prevent audible distortions of potential spectral splatter. The three bottom panels of Fig. 9 show spectrograms of the three vowels.

Table 1. Formant center frequencies and sound pressure levels of the formant combinations.

	Distractor	**/a/**	**/i/**	**/ø/**
F1 center frequency (Hz)	1000	795	294	443
F2 center frequency (Hz)	-	1301	2208	1497
F3 center frequency (Hz)	-	2795	2294	2443
F4 center frequency (Hz)	-	3795	3294	3443
Part 1: F1 + F3 (dB SPL)	65	65	65	65
Part 2: F2 + F4 (dB SPL)	-	59	51	52

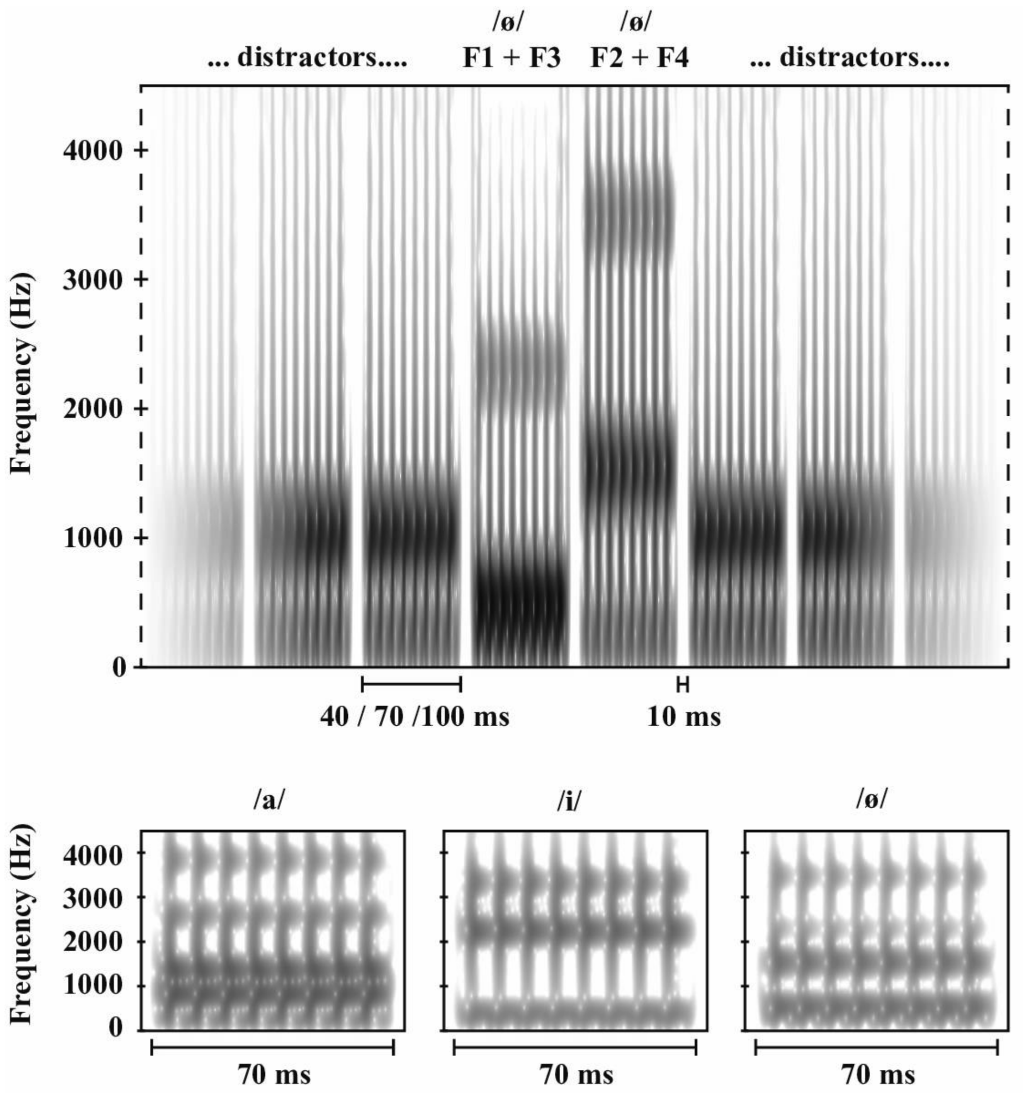


Figure 9 - Experiment 2: The top panel shows the spectrogram (window length = 5 ms; dynamic range = 70 dB) of a typical Lag 1 trial. From left to right, a number of distractor stimuli are presented, followed by part 1 (F1 + F3) of the vowel /ø/, and then part 2 (F2 + F4) of the same vowel, followed by more distractor stimuli. Stimuli were 40, 70, or 100 ms in duration, and were always separated by a 10 ms silent gap. The three bottom panels show spectrograms (window length = 5 ms; dynamic range = 45 dB) of the three 4-formant vowels /a/, /i/ and /ø/, in this example with a duration of 70 ms.

Procedure. The participants were asked to classify the targets as one of five response alternatives; the three different vowels, a tone that was lower in timbre than the distractor, or a tone that was higher than the distractor. All response alternatives were labeled on the numerical keyboard.

First, participants had to be trained to be able to identify all different targets. Therefore, they were given a few minutes to listen to each target (embedded in a short series of distractors) as often as they wanted until they felt acquainted with the targets. After that, they were given a short training session in which they were presented with the targets one by one. They then had to indicate which target they thought was presented, and they received visual feedback, together with an auditory presentation of the target they responded with and the presented target. Once the participants were able to distinguish the targets, a short final practice session followed consisting of a number of practice trials, which were similar to those in the experiment. Afterwards, the actual experiment started and consisted of 513 trials. A trial consisted of a series of 18 sequential stimuli, from which one or two could be targets and the rest distractors. On 94.74% of all trials two targets were presented, in which both targets should belong to the same formant pair (i.e., T1 as F1 and T2 as F2, or vice versa). T1 appeared as the fifth or seventh stimulus, and T2 appeared at Lag 1, 3 or 8 (each 31.58 % of all trials). On 5.26% of all trials T1 was the single target, in which it could be a vowel (1.75%) or single formant (low tones, 1.75%; high tones, 1.75%). Stimuli had durations of 40, 70 or 100 ms (1/3 of all trials each), and were separated by a 10 ms gap. The top panel of Fig.9 shows a spectrogram of a part of a typical Lag 1 trial.

The participants started a trial by pressing the spacebar. After each stream of stimuli, the participants entered what they heard as first and second targets in their perceived order. When participants only heard a single target, they were able to give an empty response as second target by pressing the Enter key. The experiment, including the training session, lasted approximately 1.5 hours for the younger adults and 2 hours for the older adults.

Data analysis. To classify a single response as a strict integration, the response should be the vowel that would have been the product of the combination of both targets. For example, if a participant reported to have only heard the /a/ and no other target, and T1 was the F1+F3 of /a/ and T2 the F2+F4 of /a/ (or vice versa), then this report would be classified as a strict integration. Otherwise, the data analysis was similar to that of Experiment 1, except that the offset for T1 accuracy was ln(54) ≈ 3,99.

Results

A full factorial analysis (WCM = exchangeable, α = 23.830) was performed on the relative frequencies of strict integration, which are shown in Fig. 10. The frequency of strict integrations was significantly affected by lag, $\chi^2(2, N = 396) = 154.9$, $p < 0.001$, by stimulus duration, $\chi^2(2, N = 396) = 51$, $p < 0.001$, and by their interaction lag*duration, $\chi^2(4, N = 396) = 32.4$, $p < 0.001$. As shown in Fig. 10, strict integrations were most frequent at Lag 1, and their frequency decreased with longer lags and longer stimulus durations. Strict integrations were also affected by group, $\chi^2(1, N = 396) = 5.3$, $p < 0.025$, as well as by the interaction of group*lag*duration, $\chi^2(3\ N = 396) = 9.3$, $p < 0.03$.

An additional analysis for Lag 1 only (WCM = autoregressive, α = 2.56) showed that stimulus duration was a significant factor, $\chi^2(2, N = 132)$ 15.3, $p < 0.001$, which indicates that shorter stimulus durations resulted in more reports of strict integrations. Also, older adults marginally reported more strict integrations at Lag 1 over all three durations, $\chi^2(1, N = 132) = 3$, $p = 0.085$. These effects can be seen in more detail in Fig. 10.

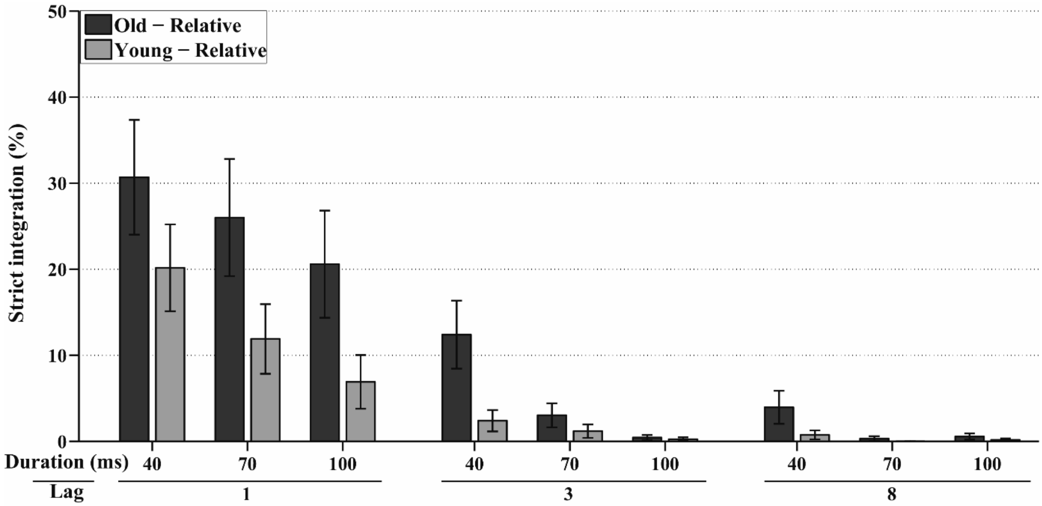


Figure 10 - Experiment 2: This graph shows the estimated marginal means of the analyses of relative frequency of strict integrations for all combinations of stimulus duration, lag, and age, as a percentage of the total number of trials in which both target identities were preserved. Error bars represent ± 1 standard error of the mean.

Another full factorial analysis (WCM = unstructured, α = 3.587) was performed on T1 accuracy, shown in Fig. 11. T1 accuracy was significantly affected by lag, $\chi^2(2, N = 396) = 74.2$, $p < 0.001$, and stimulus duration, $\chi^2(2, N = 396) = 34.5$, $p < 0.001$, as well as by their interaction lag*duration, $\chi^2(4, N = 396) = 86.5$, $p < 0.001$. Fig. 11 reveals that T1 accuracy was higher for each stimulus duration at longer lags, as well as for each lag when the stimulus durations were longer. The accuracy of T1 also differed per age group, $\chi^2(1, N = 396) = 5.6$, $p < 0.02$, indicating that the younger group performed better overall.

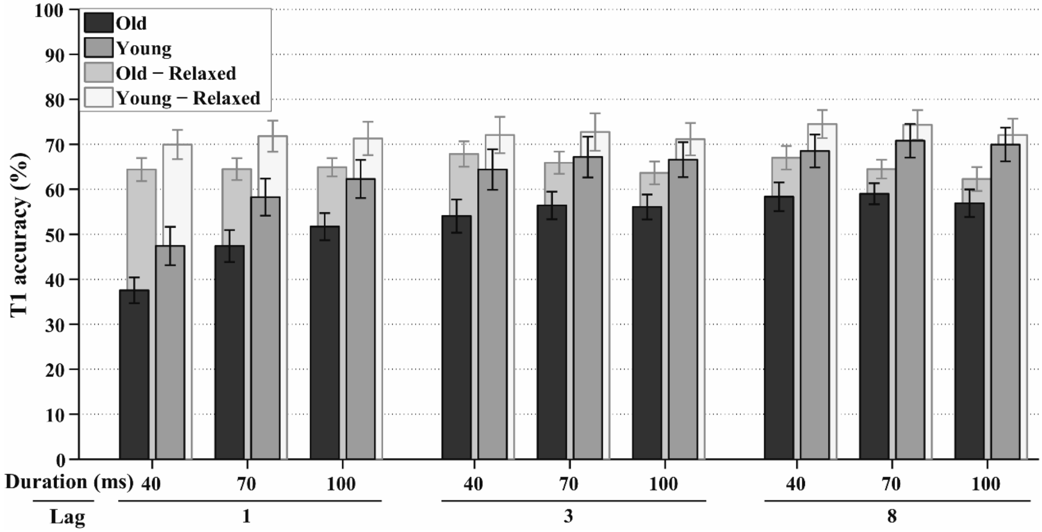


Figure 11 - Experiment 2: The solid bars at the front show the estimated marginal means of the analyses on T1 task performance in percent correct, plotted for all combinations of stimulus duration, lag, and age group. The transparent bars at the back show the same analyses if report order is ignored. Error bars represent ± 1 standard error of the mean.

The last full factorial analysis (WCM = autoregressive, α = 3.757) was performed on T2|T1 accuracy, shown in Fig. 1. T2|T1 accuracy was significantly affected by lag, $\chi^2(2, N = 396) = 17.5$, $p < 0.001$ and lag*duration, $\chi^2(4, N = 396) = 13.6$, $p < 0.01$. Fig. 12 reveals that T2|T1 accuracy was higher for each stimulus duration when lags were longer, as well as for each lag when the stimulus durations were longer (except for Lag 3 and 8 from 70 to 100 ms). The accuracy of T2|T1 also differed per age group, $\chi^2(1, N = 396) = 10.1$, $p < 0.002$, indicating that the younger adults were also better able to identify the second target.

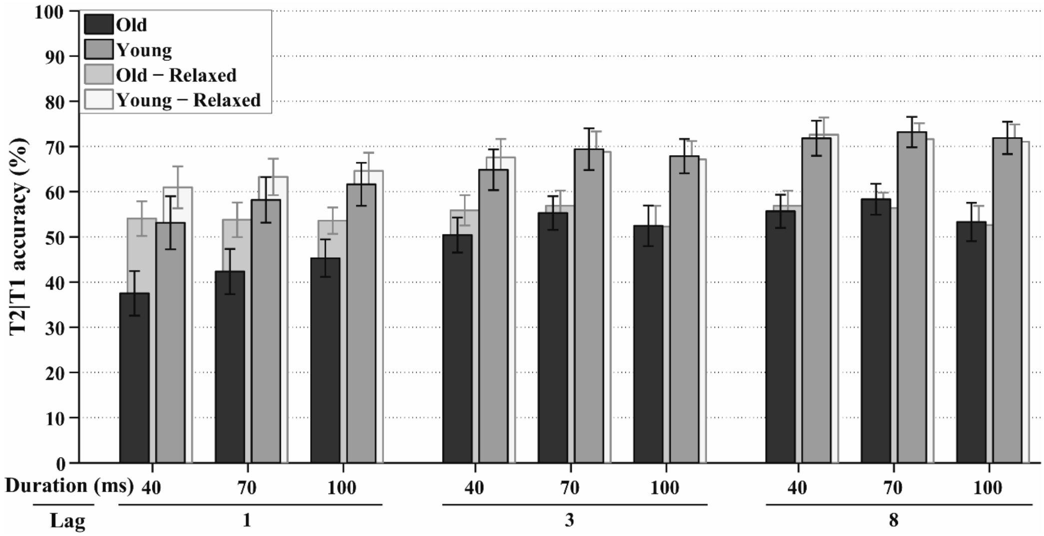


Figure 12 - Experiment 2: The solid bars at the front show the estimated marginal means of the analyses on T2|T1 task performance in percent correct, plotted for all combinations of stimulus duration, lag, and age group. The transparent bars at the back show the same analyses if report order is ignored. Error bars represent ± 1 standard error of the mean.

Comparison of Experiment 1A and Experiment 2

To analyze whether temporal integration in both rapid serial presentation experiments occurred similarly, we performed a GEE test with experiment, age group and stimulus duration as factors, on the strict integration data for Lag 1 only (WCM = unstructured; α = 6.807). The test revealed that experiment, $\chi^2(1, N = 246) = 5.4$, $p < 0.025$, age, $\chi^2(1, N = 246) = 13.2$, $p < 0.001$, and duration, $\chi^2(2, N = 246) = 64.4$, $p < 0.001$, were signifant main factors, as expected. The significant interaction effects were experiment*duration $\chi^2(2, N = 246) = 6.5$, $p < 0.04$ and age*duration $\chi^2(2, N = 246) = 16.7$, $p < 0.001$. The significant effect of experiment indicates that temporal integration was more frequent in the visual domain, as is evident from comparing Fig. 3 and Fig. 10. The interaction effect of experiment and duration indicated that integration decreased more sharply as duration increased in the visual modality. The interaction between age and duration showed that overall, this decrease was attentuated for the older participants; they integrated comparatively more at the longer durations. However, the absence of interaction effects of experiment*age and experiment*age*duration, indicate that age did not influence temporal integration differently per experiment. This means that aging affected temporal integration similarly in both modalities, even if it appeared from the individual analysis of Experiment 2 to do so less strongly in audition.

GENERAL DISCUSSION

Previous literature provided evidence that aging results in more temporal integration in vision, however, evidence for the auditory domain remained inconclusive (e.g., Horváth et al., 2007). Therefore, the primary aim of this study was to investigate if aging affects temporal integration similarly in the visual and auditory domains. To this end, we conducted two rapid serial presentation experiments, visual and auditory, aiming to obtain a direct, comparable measure of temporal integration in each modality.

The results of the visual task (Experiment 1A and 1B) showed that temporal integration was significantly affected by aging at Lag 1. The older adults reported more temporal integration overall than the younger adults did. Most importantly, the interaction effect of age and stimulus duration at Lag 1 (where both targets succeeded each other directly) was significant. This showed that for older adults visual temporal integration decreased less steeply with increasing stimulus duration, which means that the older adults integrated more at longer stimulus durations, as would be expected for a longer temporal window of integration. The results of the auditory experiment, however, showed a weaker aging effect on temporal integration: Older adults reported only marginally more temporal integrations at Lag 1 than younger adults. Also, there was no significant interaction effect between age and duration at Lag 1. Yet, the analysis of temporal integration at Lag 1 between both experiments revealed that the general pattern of performance was not reliably different. In other words, age influenced temporal integration similarly, even if temporal integration was most apparent for the visual modality as indicated by a significant main effect of experiment (cf. Fig. 3 and 10). From these facts combined we can conclude that aging does affect temporal integration in both the visual and auditory domains, but that the effect may be somewhat weaker in the latter.

Locus of age-related differences in temporal integration

In the current experiments, we aimed to minimize the differences in visual and auditory sensory properties between the age groups, so that any differences in results could be attributed to differences in cognitive rather than perceptual capabilities (cf. Lindenberger and Baltes, 1994). Because it is not feasible to fully remove all sensory differences between the age groups, we aimed to reduce the differences to a minimum by using participants that had normal vision and hearing according to the respective standards (International Council of Ophthalmology, 2002; Martini 1996). It must nonetheless be acknowledged that small sensory differences between the groups did remain, which might have contributed to differences in temporal integration. However, the results of Experiment 1B suggested that such a sensory effect can be discounted, since the data showed a pattern contrary to what would be expected if sensory degradation caused the age-related differences in temporal integration: We found less rather than more integration with reduced illuminance.

It therefore seems more likely that the differences in temporal integration originate from a more upstream locus in the perceptual processing pathway. For instance, older people have decreased early sensory memory abilities for short, individual stimuli (Fogerty, Humes and Busey, 2016), making it harder to successfully keep fine-grained, individual stimuli in store. This might result in temporally blurred representations due to longer integration windows. Older people also seem to have more difficulties with separating and encoding short, individual, sequential stimuli because of decreased temporal processing capabilities, which might again result in overlapping representations. Supporting evidence has been obtained from gap detection tasks (Di Lollo, Arnett and Kruk, 1982; Humes et al., 2009), in which younger people can detect smaller gaps, and from temporal order judgments tasks, in which older people need longer ISI and stimulus durations to successfully judge the order of two sequential visual or auditory stimuli (Kolodziejczyk and Szelag, 2008; Ulbrich et al., 2009).

Indeed, by themselves such more function-specific theories are already quite capable of explaining why older people may have longer temporal windows and integrate more than younger people. However, it may be noted that the concept of cognitive slowing arguably encompasses these more specific theories. To recap, the processing-speed theory states that cognitive slowing leads to degraded cognitive functioning (Salthouse, 1996; Madden and Allen, 2015), which impacts perception according to the common cause principle. An individual with slower cognitive processing speed can carry out fewer cognitive operations within a certain timeframe (i.e., decreased temporal processing capabilities). Consequently, with limited processing time, subsequent cognitive operations are left with less processing time as earlier operations are taking longer to finish. Because of this, memory traces of the results of earlier operations may decay or become less strong, which make them susceptible for merging with subsequent memory traces. It therefore seems most parsimonious to refer more generally to cognitive slowing as the underlying mechanism that affects temporal integration with aging, regardless of the modality.

Although a general theory for the presently observed effects is appealing, the current data leave the possibility that the prominence of time in audition can at least slightly weaken the age-related differences in that modality. However, not all alternative explanations for this slight discrepancy between modalities can be ruled out. Because sensory and cognitive aging may correlate (e.g., Humes, Busey, Craig, and Kewley-Port, 2013; Roberts and Allen, 2016; Wayne & Johnsrude, 2015), the strict exclusion criteria applied out of necessity in Experiment 2 may have resulted in a relatively high-performing sample, which may have translated into comparatively modest integration rates. Thereby, the age-related effect may have become more difficult to detect. Another possibility is that the weaker effect in audition was due to the nature of the stimuli. One might suppose that the targets in the visual experiment were less meaningful than those in the auditory experiment (i.e., vowels), and that this difference could have mediated the integration process such that auditory targets were less integrated. This account nevertheless seems problematic, because (1) not all auditory targets were meaningful vowels, (2) integrated reports could only consist of vowels combined from complex tones, which means that an increase in reports of integrated vowels should be expected, and (3) the symbols used in the visual experiment might also be regarded as meaningful (consider for instance the target “X”).

Relation to neurophysiology and attentional blink

In neurophysiological terms, age-related cognitive decline is associated with myelin loss in the white matter of brain regions that myelinate late during brain development (Lu et al., 2011; Salami et al., 2012; Lu et al., 2013), such as the prefrontal cortex (often associated with executive functioning, memory and attention) and the genu of the corpus callosum, which connects the prefrontal cortex on both hemispheres (Bloom and Hynd, 2005). Because the axons in these regions are less thickly myelinated, they are more fragile and sensitive to age-related degradation. In turn, such degradation diminishes the myelin's function to accelerate transmission speed of action potentials through leaping conduction, which could possibly lead to cognitive slowing. Because the prefrontal cortex is related to attention and working memory, a general account of cognitive slowing thus fits our results quite well. Namely, in the currently used rapid serial presentation tasks, subjects need a sufficient level of attention and working memory capacity to successfully detect, identify and remember the rapidly presented targets while ignoring intermediate distractors.

Furthermore, according to the simultaneous type serial token model (Bowman and Wyble, 2007) two targets can be combined into a single target representation or episodic memory trace when the temporal overlap between the activation of both targets is adequate. Perceptually combining two targets in such a way costs less mental effort, as was shown by

Wolff et al. (2015), meaning that working memory is burdened less. Because older adults generally struggle more on attentional and cognitive tasks (Craik and Salthouse, 2011), it is conceivable that they use this temporal integration mechanism more frequently, as it may serve as a compensation mechanism to save mental resources. Most compensation mechanisms that are used by older adults result in increased brain activity compared to younger adults, to compensate for the age-related changes in the brain (Grady, 2012). In our tasks, to successfully detect, identify, remember and keep up with the rapidly presented targets and ignore distractors, it is conceivable that older adults integrate more because they have less mental resources or neuronal connections to perform this demanding task.

If so, it might be hypothesized that the brain activity of older adults in the prefrontal cortex increases as a way to keep up with the fast pace, resulting in an attempt to increase attention to the targets. Previous research showed that if more attention is given to targets temporal integration also increases (Visser and Enns, 2001), and also that successful temporal integration is related to increased amplitudes of the N1, N2 and late P3, which are event-related potential components related to attention (Akyürek, Schubö and Hommel, 2010). Note that even though temporal integration might come with increased brain activity, it is nonetheless less demanding (or costs less mental effort) than keeping up with each single stimulus at a time, making it a suitable compensation mechanism for older adults with fewer neuronal connections (and thereby likely fewer mental resources) to begin with. In practice, such compensation would result in a prolonged temporal integration window, as longer periods are covered in a single episodic memory trace, which can be seen in our experiments where the older adults integrated more at longer durations.

These interpretations fit well with previous attentional blink (AB) results. The AB is expressed in the difficulty of perceiving the second of two targets (typically in RSVP), if it arrives between 150-500 ms after the first (Raymond, Shapiro, & Arnell, 1992). Importantly, recent work on individual differences suggests that people with a larger AB tend to integrate more (Willems, Saija, Akyürek and Martens, 2016), which is in line with task performance in terms of effective allocation of cognitive resources, as given above. Furthermore, previous research has also shown that the AB is larger for older adults in both modalities (Lahar, Isaak and McArthur, 2001; Slawinski and Goddard, 2001). The current results show a similar pattern, both for integration, as discussed, and for target accuracy also: Age had a significant effect on T1 and T2|T1 accuracy in both modalities, meaning that for both measures older adults had lower accuracy over all conditions. One caveat is that even though we controlled for normal visual and auditory acuity, in practice the acuity was on average slightly better for the younger groups, which might have contributed to the differences in accuracy.

Finally, a further advantage of a prolonged temporal integration window, besides the reduction of mental effort, is that it might be beneficial for high-level compensatory mechanisms for better perception of degraded speech, such as measured in studies of the phonemic restoration effect (Warren, 1970; Başkent, 2012). With phonemic restoration, listeners are able to restore degraded speech that contains missing speech segments that are filled by loud noise, by using top-down knowledge to fill in the missing segments and combine the available and filled-in loose segments into coherent understandable speech. Saija et al. (2014b) showed that older adults, in some conditions, have a larger restoration effect than younger adults, and concluded that this might be due to the older adults' superior language skills, vocabulary and world knowledge. However, in light of the current results, it might be that temporal integration plays a role as well. Namely, Fig. 10 shows that with auditory stimuli older adults integrated more at Lag 3 than younger adults (most prominent at 40 ms stimulus duration, and similar to the visual task). Normally, temporal integration would occur when two targets are presented in succession without intermediate distractors. However, for the older adults in this case, integration also happened with intermediate distractors at Lag 3. Such integration of two targets spanning over two intermediate distractors is not seen with young adults. With phonemic restoration, listeners also have to combine information of speech segments that are separated or masked by intermediate noise. Therefore it is conceivable that a prolonged temporal integration window, as is seen with older adults, might have a positive effect on the phonemic restoration ability.

Conclusion

In summary, the current results show that the older adults integrated overall more than the young adults, independent of modality. The older adults also integrated comparatively more at longer durations than the young adults. This effect was most clearly observed in the visual domain, and seemed less pronounced in audition. These results seem to reflect a general, cognitive-perceptual change with age, with the tentative addition that the prominence of time in audition may weaken this effect for auditory temporal integration.

CHAPTER 4

Perceptual restoration of degraded speech is preserved with advancing age

J. D. Saija, E. G. Akyürek , T. C. Andringa and D. Başkent

Published in Journal of the Association for Research in Otolaryngology (2014), 15(1), 139–148. doi: 10.1007/s10162-013-0422-z.

ABSTRACT

Cognitive skills, such as processing speed, memory functioning, and the ability to divide attention, are known to diminish with aging. The present study shows that, despite these changes, older adults can successfully compensate for degradations in speech perception. Critically, the older participants of this study were not pre-selected for high performance on cognitive tasks, but only screened for normal hearing. We measured the compensation for speech degradation using phonemic restoration, where intelligibility of degraded speech is enhanced using top-down repair mechanisms. Linguistic knowledge, Gestalt principles of perception, and expectations based on situational and linguistic context are used to effectively fill in the inaudible masked speech portions. A positive compensation effect was previously observed only with young normal hearing people, but not with older hearing-impaired populations, leaving the question whether the lack of compensation was due to aging or due to age-related hearing problems. Older participants in the present study showed poorer intelligibility of degraded speech than the younger group, as expected from previous reports of aging effects. However, in conditions that induce top-down restoration, a robust compensation was observed. Speech perception by the older group was enhanced, and the enhancement effect was similar to that observed with the younger group. This effect was even stronger with slowed-down speech, which gives more time for cognitive processing. Based on previous research, the likely explanations for these observations are that older adults can overcome age-related cognitive deterioration by relying on linguistic skills and vocabulary that they have accumulated over their lifetime. Alternatively, or simultaneously, they may use different cerebral activation patterns or exert more mental effort. This positive finding on top-down restoration skills by the older individuals suggests that new cognitive training methods can teach older adults to effectively use compensatory mechanisms to cope with the complex listening environments of everyday life.

INTRODUCTION

Age-related decline in sensory and cognitive capabilities, more specifically of fluid intelligence, can affect the quality of life negatively (Dalton et al. 2003). On the sensory side, visual and auditory capabilities (Swenor et al. 2013), and on the cognitive side, cognitive processing speed (Salthouse 1996; Kennedy and Raz 2009), short- and long-term memory functioning (Kemper et al. 2003; Charlton et al. 2010; Tonoki and Davis 2012), and top-down suppression (Gazzaley et al. 2008; Wild-Wall and Falkenstein 2010; Janse 2012) decline with increasing age. The present study shows that older adults have mechanisms to effectively compensate for the negative age-related changes to enhance perception, perhaps by relying more on the crystallized intelligence that can stay the same or become better over a lifetime (Cattell 1971; Baltes 1993).

To show such compensation, the phonemic restoration paradigm was used in this study (Warren 1970; Powers and Wilcox 1977; Verschuure and Brocaar 1983; Bashford et al. 1992). Here, missing parts of speech are perceptually filled in by the listener, which can improve speech intelligibility in noisy environments. Such restoration depends on an interaction between knowledge-driven and signal-driven processes (Başkent 2012), where linguistic skills, world knowledge, experience, expectations, situational context and Gestalt rules of perceptual grouping are used to form perceptual hypotheses based on sentential context and semantic, syntactic, spectral and temporal cues from the speech signal (Samuel 1981; Bashford et al. 1992; Srinivasan and Wang 2005; McDermott and Oxenham 2008; Riecke et al. 2009; Groppe et al. 2010). Degraded speech, in which parts of the signal are missing, reduces speech redundancy and increases reliance on the cognitive and linguistic mechanisms for top-down repair (Stenfelt and Rönnberg 2009).

A positive compensatory effect of phonemic restoration has previously only been demonstrated with young normal hearing adults (Başkent et al. 2009; Başkent 2012; Benard and Başkent 2013b), but the effect diminished with younger adults tested with simulated hearing devices and older adults with hearing loss (Başkent et al. 2010; Başkent 2010;

Başkent 2012). Hence, the question remained whether older individuals can benefit from the restoration mechanisms using their knowledge-based linguistic skills. In the present study, by presenting speech at varying levels of degradation, it was explored if, and when, older listeners can engage these compensatory restoration mechanisms. Additionally, as a way of inspecting the effects from age-related cognitive slowing, restoration was measured at varying speech rates. By selecting older participants with normal hearing, potential effects of age–related sensory decline were eliminated (Hoffman et al. 2012), which enabled us to focus on cognitive aspects. Specifically no further pre-selection was used, to have a representative group of older listeners that was not trained nor high-performing per se.

METHODS

Participants

Appropriate selection of the older group was crucial for the study to ensure the results were mainly caused by aging, and not contaminated by age-related hearing loss and the resulting limitations in audibility. The older participants were selected to be of sufficiently advanced age to show age effects on speech perception (Mahncke et al. 2006; He et al. 2008; Wong et al. 2009; Adams et al. 2012; Füllgrabe 2013) while displaying nearly normal hearing (Hoffman et al. 2012). As a result, from the 21 older participants who applied for the study and self-reported to have normal hearing, 12 (57%) qualified to have normal hearing, as defined by our inclusion criterion, during the screening.

Twelve young (six female, six male, mean age 22 years, ranging from 19 to 26 years) and twelve older (four female, eight male, mean age 66 years, ranging from 62 to 77 years) native Dutch speakers participated in this study. All participants were tested for normal hearing (defined as four-tone pure-tone average of the thresholds measured at the audiometric frequencies of 500, 1000, 2000 and 4000 Hz less than or equal to 20 dB HL for the better ear, based on Stephens (1996); see Fig. 1 for individual hearing thresholds). Participants further reported no hearing problems and no language disorders (e.g. dyslexia). They were naive to the testing materials and methods, and unaware of the purpose of the study. Written informed consent was obtained before participation and the study was approved by the Ethics Committee of the University of Groningen, Department of Psychology. Participation was financially compensated.

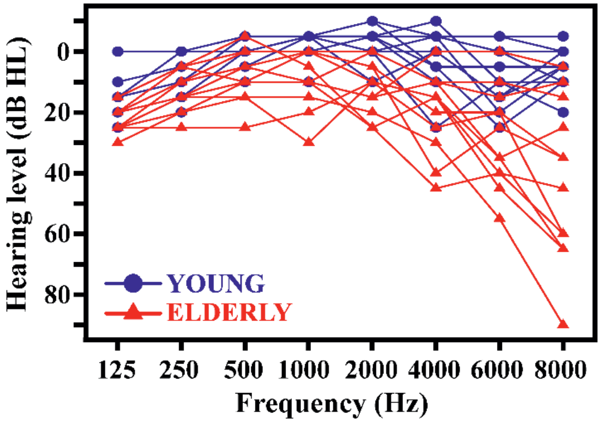


Fig. 1 Individual hearing thresholds shown for the better ear for young and older adults

Apparatus

The experiment was programmed in Matlab (version 7.10.0.499, 32-bit) and run under Mac OS X (10.5.8) on a Mac Pro. The audio stimuli were diotically presented through a Sennheiser HD 600 headphone, connected to an Echo Audiofire 4 external soundcard SPDF output and a Lavry Engineering DA10 digital-to-analog converter. Participants were seated in a sound-isolated testing booth. Participant responses were recorded with an Alesis Palm Track digital sound recorder in MP3 format for offline double-checking of the scores.

Stimuli

The speech corpus consisted of meaningful and everyday Dutch sentences sampled at 44.1 kHz (Versfeld et al. 2000). Originally, the authors created a subset of 78 lists (39 lists spoken by one male speaker and 39 spoken by one female speaker) of 13 sentences each, with the purpose of measuring speech reception thresholds in stationary speech-shaped noise. Each sentence in the corpus is syntactically and grammatically correct. Each sentence consists of four to nine words and each word consists of three or fewer syllables. In the present study, of the original 39 balanced lists of 13 sentences, spoken by the male talker, 33 lists were used. The presentation order was randomized for each participant.

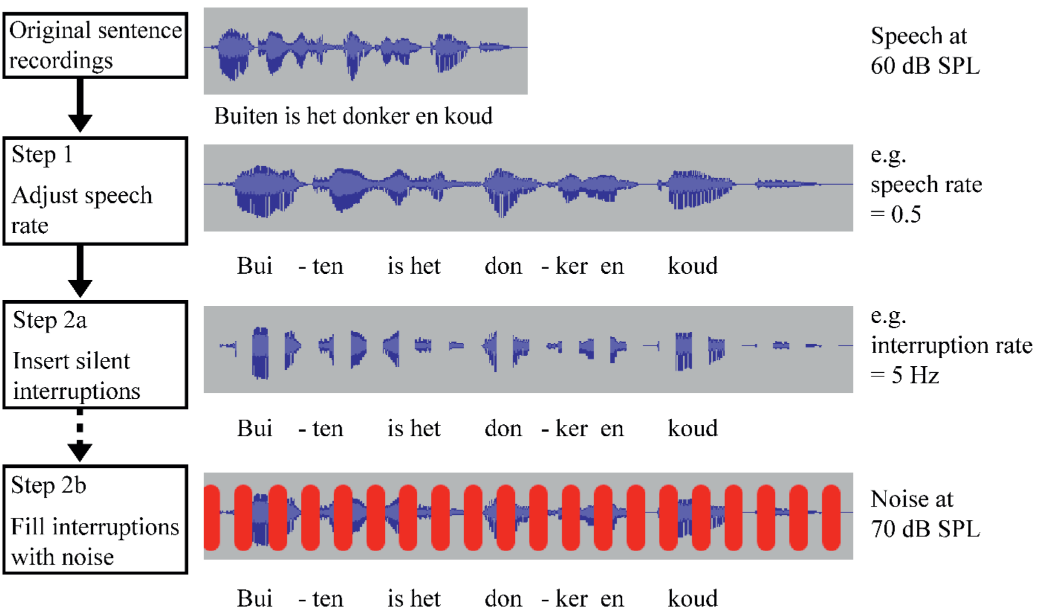


Fig. 2 Schematic representation of stimulus construction. Blue and red lines represent speech and filler noise, respectively. At step 1, the speech rate of the original sentence recordings was altered, by slowing down (speech rate = 0.5) or speeding up (speech rate = 2) with a factor of 2. At step 2a, periodic silent interruptions were inserted in the sentences

at various interruption rates (0.625, 1.25, 2.5, 5, 10 and 20 Hz). At Step 2b, the silent gaps were filled with the noise bursts

Fig. 2 schematically shows the steps that were involved in creating the speech stimuli. The first step was to change the speech rate (SR; 0.5x, 1x and 2x the original speech rate) by compressing or expanding the sentence recordings without altering the voicing pitch. For this purpose, the pitch-synchronous-overlap-add (PSOLA) method (Moulines and Charpentier 1990) was employed using PRAAT (version 5.3.12) software (Boersma 2002), using the default settings (time steps of 10 ms, minimum pitch of 75 Hz and maximum pitch of 600 Hz). The second step was to introduce periodic gaps at varying interruption rates (logarithmic steps; 0.625, 1.25, 2.5, 5, 10 and 20 Hz), by modulating the time-modified sentences with a periodic square-wave signal. The square-wave period depended on the interruption rate, e.g. an interruption rate of 5 Hz produced a period of 200 ms, with an ON- and OFF-phase, both with an equal length of 100 ms (duty cycle of 50%). A raised cosine ramp of 5 ms was placed at each onset and offset to prevent audible distortions due to spectral splatter.

To measure the phonemic restoration benefit, in half of the conditions, noise was used to fill the silent gaps. With the filler noise, the interrupted speech stream is more likely to be perceived as a continuous perceptual object, but simultaneously an ambiguity is introduced to the brain. It cannot tell if parts of speech are indeed missing or if they are simply masked. This seems to induce perceptual grouping mechanisms, helping the brain to form an object from audible speech samples and thereby perceptually filling-in for missing speech (Shahin et al. 2009). Hence, the noise-filled conditions induce top-down repair mechanisms, and the improved intelligibility resulting from adding noise to silent gaps is taken as a measure of perceptual restoration benefit. The filler noise was a speech-shaped steady noise created by averaging the spectra of the male speech stimuli and randomizing the phase of the average spectrum (Versfeld et al. 2000). By inverting the phase of the same square-wave modulating function and applying it to the filler noise, the filler noise bursts were produced. These bursts were added to the interrupted speech and filled the silent gaps, which resulted in speech interrupted by filler noise.

In all conditions, with or without the filler noise, the speech presentation level was fixed at 60 dB SPL, and the filler noise presentation level at 70 dB SPL, producing a signal-to-noise ratio of -10 dB SPL (Powers and Wilcox 1977; Başkent 2010; Başkent 2012).

Experimental conditions

To ensure that the effects observed in the experiment were not caused by low baseline scores due to speech rate manipulation or some other age-related factors, a baseline performance was measured at three speech rates (speech rate = 0.5, 1, 2) without any interruptions in sentences. All subjects of the study had a baseline score close to ceiling performance (Fig. 3). Next, the interruption conditions (0.625, 1.25, 2.5, 5, 10 and 20 Hz) were applied to the three speech rates. For slow speech, 10 and 20 Hz were excluded, and for normal and fast speech 0.625 Hz, to have speech segment durations more comparable across different conditions (based on a pilot study). All 14 conditions were tested twice, once with and once without the filler noise, producing 31 trials of 13 sentences each, including the three baseline measurements.

Procedure

Each condition was tested with one unique, randomly selected sentence list. Before the presentation of the first test sentence, the same introductory sentence ‘Buiten is het donker en koud’ (‘Outside it is dark and cold’) was played to indicate the beginning of a new list. Because the introductory sentence was processed the same way as the manipulations of the particular condition it also served to prime the participants for the specific manipulation that was about to be tested. To indicate the start of each sentence, participants heard a short beep preceding the stimulus.

After listening to each stimulus, the task of the participants was to reconstruct the sentences and formulate them into meaningful, correct Dutch sentences and to verbally report these. Guessing was encouraged to ensure that participants could report what they thought they heard even when they were not sure. Participants could also report only parts of sentences when they were unable to create entire meaningful sentences. Scoring was done online by the experimenter (first author), who sat outside the testing booth and listened to the participants' responses through a headphone via the digital audio recorder. An annotation program developed in Matlab (version 7.10.0.499, 32-bit) was used. A Matlab GUI showed the list number and sentence number and could advance to the next sentence when a participant finished reporting what he / she heard. The experimenter was unaware of which experimental condition was being tested. After each condition / list the program would automatically calculate the correct percentage of annotated words with respect to the total words of all sentences in the list used in a particular trial. Each session was also recorded with the digital audio recorder for offline annotation to double check potential errors in the online annotation. The annotation rules were in line with the rules mentioned by Başkent (2012).

After the screening and briefing of participants, and following the baseline measurements (randomly assigned to lists 1-3), a short training session without feedback of one condition (interruption rate of 5 Hz and speech rate of 1) with and without filler noise was provided, with each randomly assigned to lists 32 and 33. After training, each one of the 28 interruption conditions was tested with a randomly assigned list from lists 4-31. All participants completed the experiment in a single session, which lasted on average one and a half hours, including the initial screening and occasional breaks.

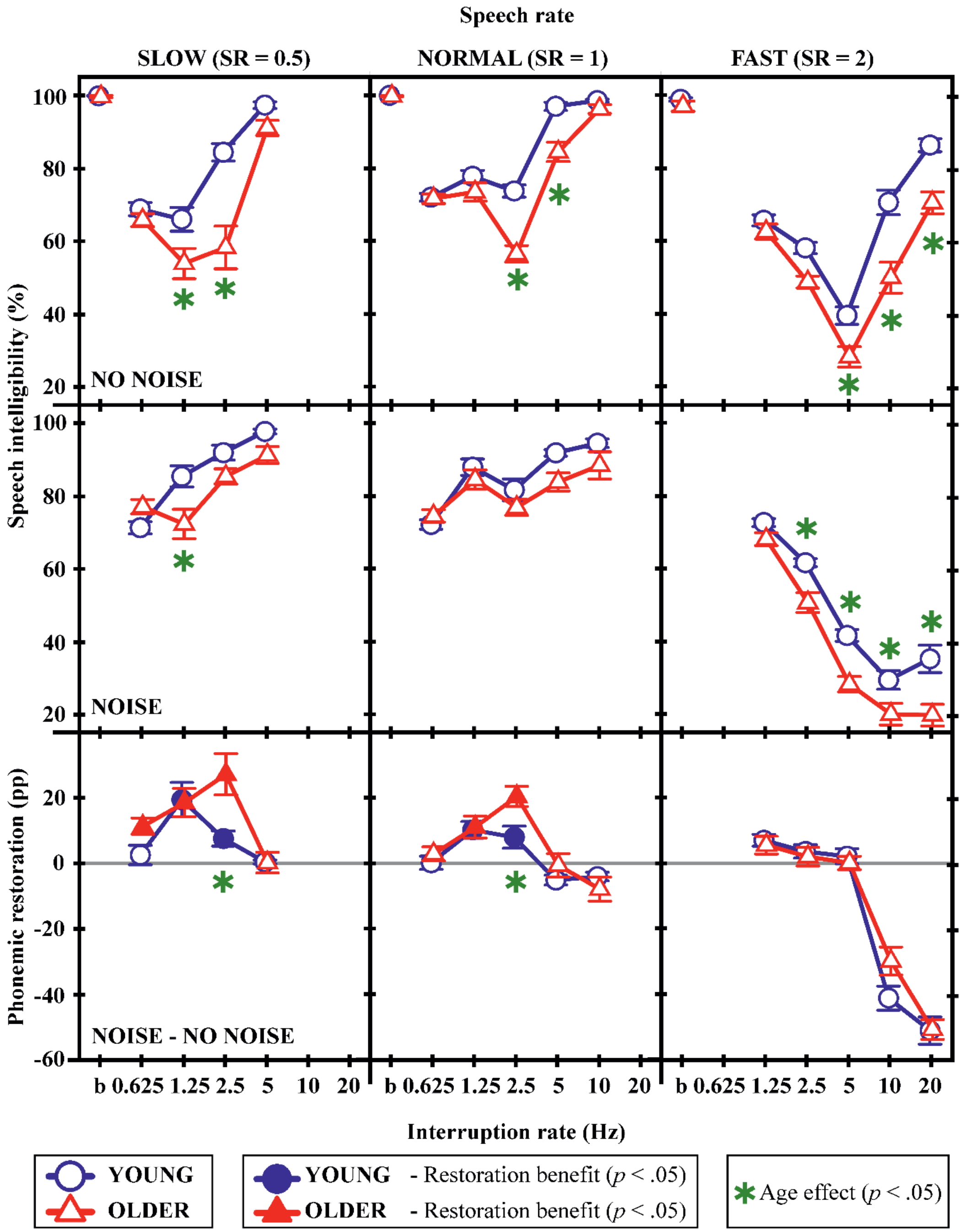

Speech rate
SLOW (SR = 0.5)
NORMAL (SR = 1)
FAST (SR = 2)
Speech intelligibility (%)
NO NOISE
NOISE
Phonemic restoration (pp)
NOISE - NO NOISE
b 0.625 1.25 2.5 5 10 20
Interruption rate (Hz)
YOUNG
OLDER
YOUNG - Restoration benefit (p < .05)
OLDER - Restoration benefit (p < .05)
Age effect (p < .05)

Fig. 3 Intelligibility scores and restoration benefit shown for young and older groups and per speech rate. Speech intelligibility scores, averaged for each age group, are shown for interrupted speech with silent gaps (top panels) and with filler noise (middle panels). The lowest panels directly show the restoration effect in percentage points (pp), calculated by taking the difference in scores from top and middle rows. In the lowest panels, all values above 0, shown by the solid, horizontal gray lines, denote a restoration benefit, and if the benefit was significant, it is further marked by a filled symbol. Results with slow, normal and fast speech rates are shown in the left, middle and right columns, respectively. SR stands for speech rate. In all panels, the age effect was tested with post hoc tests, and the significant effects are denoted by "*" for the corresponding conditions. The leftmost symbols in the upper panels (indicated with the letters 'b' on the x-axes) show the baseline performances with uninterrupted original sentences, which were identical between the age groups. This made sure that the difference in data between the young and older groups was indeed due to experimental manipulations, and not some other inherent age-related factor. Error bars show ±1 standard error

Fig. 3 shows the average scores for intelligibility of interrupted speech with silent gaps (top panels), for interrupted speech combined with filler noise (middle panels), and for the compensatory restoration benefit directly (lower panels) per subject group. The columns from left to right show the effect of changing the speech rate from slow to fast. Firstly[1], for each age group and speech rate, it was determined whether the addition of noise increased intelligibility (i.e. the restoration benefit) by performing separate RM-ANOVAs with the within-subjects factor of interruption rate and the factor that represents the addition of noise (Table 1). Secondly, to determine the age effect on restoration and overall intelligibility for each speech rate (all nine panels of Fig. 3), the data in each panel was analyzed with a separate RM-ANOVA with the between-subjects factor of age and within-subjects factor of interruption rate (Table 2). Thirdly, to determine the effect of changing speech rate on

[1] In this study, experimental conditions were chosen such that they were comparable with respect to having equal amounts of speech information . This allows one to analyze what the effect is of changing the speech rate, while keeping the amount of speech information intact (this analysis is not included in the manuscript). Practically, each speech rate manipulation needs a manipulation of interruption rate of the same magnitude to obtain equal amount of speech information. This resulting asymmetrical design prevented the implementation of an overall RM-ANOVA including all factors, because not all interruption rates were measured for all speech rates.

restoration, per age group, RM-ANOVAs were performed with within-subject factors of speech rate and the interruption rates that overlapped between speech rates (Table 3). For each RM-ANOVA, sphericity was tested with Mauchly's Test of Sphericity, and when sphericity was not assumed, degrees of freedom were adjusted using the Greenhouse-Geisser epsilon correction. Age effects were examined in more detail using Tukey's HSD post hoc tests.

Table 1. Results shown from separate RM-ANOVAs conducted to determine restoration benefit per age group and for each of the three different speech rates. Bold p-values indicate significance below alpha of 0.05. The main factor NOISE denotes the difference in intelligibility after combining the interrupted sentences with filler noise, and hence shows if there is a significant restoration effect. IR and SR stand for interruption rate and speech rate, respectively.

Table 1

Statistical analyses for restoration benefit per age group

		Speech Rate (SR)											
		Slow (SR = 0.5)				**Normal (SR = 1)**				**Fast (SR = 2)**			
Age	Source	*df*	*MSE*	*F*	*p*	*df*	*MSE*	*F*	*p*	*df*	*MSE*	*F*	*p*
Young adults	NOISE	1, 11	123.86	10.54	**0.008**	1, 11	32.74	2.73	0.127	1, 11	40.07	131.70	**0.000**
	IR	3, 33	38.75	94.81	**0.000**	2.18, 23.99	47.61	103.13	**0.000**	4, 44	61.20	47.61	**0.000**
	NOISE × IR	3, 33	42.41	10.26	**0.000**	4, 44	30.86	9.41	**0.000**	2.36, 25.98	93.13	84.88	**0.000**
Older adults	NOISE	1, 11	153.30	31.53	**0.000**	1, 11	75.47	10.22	**0.009**	1, 11	45.72	141.82	**0.000**
	IR	3, 33	91.94	37.01	**0.000**	2.23, 24.58	105.00	40.12	**0.000**	4, 44	65.61	74.94	**0.000**
	NOISE × IR	3, 33	98.17	7.97	**0.000**	4, 44	58.20	12.26	**0.000**	4, 44	59.16	61.83	**0.000**

Table 2. Results shown from separate RM-ANOVAs conducted for intelligibility of interrupted sentences (with silent intervals or filler noise) and restoration benefit, for the three different speech rates (see Fig. 2). Bold p-values indicate significance below alpha of 0.05. SR stands for speech rate. AGE and IR represent the main factors of age and interruption rate, respectively.

Table 2

Statistical analyses for age effects on overall intelligibility of interrupted sentences and the restoration benefit

		Speech Rate (SR)											
		Slow (SR = 0.5)				**Normal (SR = 1)**				**Fast (SR = 2)**			
Noise condition	Source	*df*	*MSE*	*F*	*p*	*df*	*MSE*	*F*	*p*	*df*	*MSE*	*F*	*p*
Silent intervals	Between subjects												
	AGE	1, 22	232.43	14.61	**0.001**	1, 22	63.26	25.56	**0.000**	1, 22	115.49	37.80	**0.000**
	Within subjects												
	IR	2.18, 47.92	114.13	63.44	**0.000**	2.67, 58.83	43.93	148.57	**0.000**	4, 88	70.74	90.37	**0.000**
	IR × AGE	2.18, 47.92	114.13	7.58	**0.001**	2.67, 58,83	43.93	10.72	**0.000**	4, 88	70.74	3.76	**0.007**
With filler noise	Between subjects												
	AGE	1, 22	115.86	5.34	**0.031**	1, 22	84.85	5.65	**0.027**	1, 22	151.33	22.97	**0.000**
	Within subjects												
	IR	3, 66	52.76	38.62	**0.000**	4, 88	57.47	21.99	**0.000**	4, 88	49.73	188.85	**0.000**
	IR × AGE	3, 66	52.76	7.03	**0.000**	4, 88	57.47	1.62	0.177	4, 88	49.73	2.09	0.089
Restoration benefit	Between subjects												
	AGE	1, 22	277.15	4.02	0.057	1, 22	108.21	3.10	0.092	1, 22	85.79	0.59	0.450
	Within subjects												
	IR	3, 66	140.57	13.37	**0.000**	4, 88	89.07	20.15	**0.000**	4, 88	114.14	144.18	**0.000**
	IR × AGE	3, 66	140.57	3.95	**0.012**	4, 88	89.07	2.40	0.056	4, 88	114.14	1.69	0.161

At normal speech rate, regarding the restoration benefit per age group, RM-ANOVAs indicated significant restoration only for the older group (p = .009; lower middle panel of Fig. 3 and middle column of Table 1). Nonetheless, post hoc tests showed that both age groups displayed significant restoration benefit at slow interruption rates of 1.25 (p < .001 for the young and p = .009 for the older group) and 2.5 Hz (p = .01 for the young and p < .001 for the older group). There was a significant main effect of age on overall intelligibility for both versions of the interrupted speech, with silent gaps (p < .001) and with filler noise (p = .027) (upper and middle panels of middle column of Fig. 3 and Table 2). This indicates that intelligibility of interrupted speech in general was significantly lower for the older group. Particularly with speech with silent gaps, young adults outperformed the older adults at 2.5 (p < .001) and 5 Hz (p = .002). Regarding the effect of age on restoration, there was no significant main effect for age (lower middle panel of Fig. 3 and Table 2). Hence, with normal speech rate in general, the older adults obtained comparable restoration benefit to the young adults. Interestingly, the benefit obtained by the older group was significantly larger than that of the younger group at 2.5 Hz (p = .015).

With slowed-down speech, both young (p = 0.008) and older adults (p < 0.001) showed significant restoration benefit (lower left panel of Fig. 3 and left column of Table 1). More precisely, young adults showed significant restoration at 1.25 (p < .001) and 2.5 Hz (p = .040), and older adults at 0.625 (p = .043), 1.25 (p < .001) and 2.5 Hz (p < .001). Similar to normal-rate speech, the main effects for age showed that overall intelligibility was lower for older than young adults for interrupted speech with silent gaps (p = .001) and with filler noise (p = .031) (Fig. 3 and Table 2, upper and middle panels of left column). Young adults outperformed the older adults with speech with silent gaps at 1.25 (p = .037) and 2.5 Hz (p < .001) and with filler noise at 1.25 Hz (p < .001). Further, regarding the age effect on restoration, there was no significant main effect for age (p = .057; lower left panel of Fig. 3 and Table 2). Post hoc tests indicated that the older adults obtained more benefit than young adults at 2.5 Hz (p < .001). Slowing down speech increased restoration benefit significantly only for the older adults (p = .038; lower left panel of Table 3).

With speeded-up speech, instead of a positive effect from restoration benefit, there was a negative effect of adding noise (lower right panel of Fig. 3 and right column of Table 1). Similar to slow- and normal-rate speech, overall intelligibility was higher for young adults for interrupted speech with both silent gaps ($p < .001$) and filler noise ($p < .001$) (upper and middle panels of right column of Fig. 3 and Table 2). Without noise, young adults outperformed the older adults specifically at 5 ($p = .012$), 10 ($p < .001$) and 20 Hz ($p < .001$), and with noise at 2.5 ($p = .002$), 5 ($p < .001$), 10 ($p = .013$) and 20 Hz ($p < .001$). There was no age effect on restoration.

Table 3. Results are shown from separate RM-ANOVAs conducted for investigating the effect of changing the speech rate on restoration benefit per age group. For an even comparison, these models included only interruption rates that overlapped between different speech rates, thereby excluding 10 and 20 Hz while comparing slow and normal rates, and 0.625 Hz while comparing normal and fast rates. Bold p-values indicate significance below alpha of 0.05. SR stands for speech rate and IR for interruption rate.

Table 3

Statistical analyses for speech rate effects on restoration benefit per age group

		Speech Rate (SR)							
		Restoration comparison for slow – normal				**Restoration comparison for normal – fast**			
Age	Source	*df*	*MSE*	*F*	*p*	*df*	*MSE*	*F*	*p*
Young adults	SR	1, 11	124.53	3.37	0.094	1, 11	49.14	42.61	**0.000**
	IR	3, 33	61.22	22.34	**0.000**	1.81, 19.89	130.88	60.62	**0.000**
	SR × IR	3, 33	86.96	1.19	0.327	3, 33	59.54	37.35	**0.000**
Older adults	SR	1, 11	148.24	5.56	**0.038**	1, 11	86.93	35.29	**0.000**
	IR	3, 33	169.64	15.00	**0.000**	3, 33	93.06	47.36	**0.000**
	SR × IR	3, 33	140.88	0.51	0.681	3, 33	157.13	4.37	**0.011**

DISCUSSION

Overall, the results show that older adults can benefit from the top-down repair mechanisms involved in phonemic restoration in a robust way, similar to young adults. This finding is a nice surprise, given that the intelligibility of interrupted speech, especially with silent gaps, was worse for older than younger listeners (Fig. 3 and Table 2, top and middle rows), an observation in line with previous studies with similar speech manipulations (Bergman et al. 1976; Gordon-Salant and Fitzgibbons 1993). Despite lower scores for interrupted speech overall, the relative improvement due to restoration was comparable to (and sometimes even better than) that of the younger group. Hence, the phonemic restoration ability is preserved with advanced age despite potential cognitive deterioration of fluid intelligence associated with old age that could have worked against it (Salthouse 1996; Kemper et al. 2003; Humes et al. 2006; Gazzaley et al. 2008; Kennedy and Raz 2009; Charlton et al. 2010; Wild-Wall and Falkenstein 2010; Janse 2012; Tonoki and Davis 2012; Swenor et al. 2013). This shows that older people seem to be able to compensate for the age-related decrement in perception of degraded speech, probably by relying on their gained knowledge and experiences of language (Cattell 1971; Baltes 1993; Park et al. 2002; Salthouse 2004; Pichora-Fuller 2008).

Previously, poorer restoration was observed with older listeners with hearing loss (Başkent et al. 2010; Başkent 2010). Our new findings imply that the reduced restoration must have been mainly due to hearing impairment, a sensory factor that was eliminated in this study, and not age per se. Supporting this idea, even with young participants, compensatory restoration also diminished with young participants who were tested with simulations of hearing devices (Başkent et al. 2009; Başkent 2012; Benard and Başkent 2013a; Bhargava et al. 2013). Specifically, when the bottom-up speech signal lacks the appropriate speech features that can induce top-down linguistic processes, such as lexical activation, inserting noise in the silent gaps may make the speech sound more continuous, but may not necessarily increase the intelligibility (Miller and Licklider 1950; Bhargava et al. 2013). These observations combined indicate that the top-down restoration processes depend on the state of bottom-up speech signals, in line with speech perception models that emphasize the interactive nature of the bottom-up and top-down processes for speech perception (Wingfield et al. 2005; Davis and Johnsrude 2007; Sheldon et al. 2008; Stenfelt and Rönnberg 2009; Sohoglu et al. 2012; Başkent 2012; Goy et al. 2013). In short, based on previous literature regarding phonemic restoration and hearing loss (Başkent et al. 2009; Başkent et al. 2010; Başkent 2010; Başkent 2012), it might be that even though older adults obtain lower intelligibility than young adults for interrupted speech in general, they nonetheless benefit equally from restoration if the speech signal is not further degraded as would be the case with hearing impairment or hearing devices.

In this study, an additional manipulation, namely the altering of speech rates, was introduced. This was done to further explore potential effects of age-related cognitive slowing (Salthouse 1996). Slowing down speech increased the restoration benefit by the older adults, while speeding it up made the benefit disappear for both participant groups. This observation further confirmed that processing speed plays an important role for understanding degraded speech and phonemic restoration. Slowed speech seems to give the older adults more time to process noisy speech and use available cues from the speech signal more effectively.

Previously, compensation effects have usually been shown with selected high-performing older adults (Buckner 2004; Hedden and Gabrieli 2004). For this study, no such pre-selection or testing of cognitive skills of participants occurred other than screening for normal hearing. This study shows that older listeners can use top-down restoration to enhance intelligibility of degraded speech. Based on previous studies, it suggests that older people likely use supportive, sentential context better (Pichora-Fuller 2008), effectively utilizing their lifelong experience with language and accrued word knowledge (Cattell 1971; Baltes 1993; Park et al. 2002; Salthouse 2004). Our findings are also in line with the idea that older people may use central cognitive functions differently (Reuter-Lorenz 2002; Buckner 2004; Hedden and Gabrieli 2004; Wong et al. 2009; Grady 2012) or exert more mental effort (McCoy et al. 2005;

Getzmann and Falkenstein 2011). Nonetheless, further research can help to identify the precise factors that help older people to compensate, which in turn could lead to new training methods that can help them to learn to perform better. Benard and Başkent (Benard and Başkent 2013a; Benard and Başkent 2013b) showed that training improved perception of interrupted speech, indicating that people are able to learn to use the top-down repair mechanisms more effectively. Hence, directed cognitive training might help older adults to overcome cognitive deficits in old age (Mahncke et al. 2006; Anderson et al. 2013) and better cope with the complex listening environments of everyday life.

CHAPTER 5

General discussion

In this thesis we investigated several aspects of temporal integration, comprising both the visual and the auditory modalities, as well as potential aging effects. In the introductory chapter, four research questions were posed, which were addressed in the subsequent chapters. Here I will summarize and contextualize the main outcomes.

The first question we sought to answer was whether (merging) temporal integration behaves similar in audition as it does in vision. In the second chapter we discussed previous evidence of similarities and dissimilarities in vision and audition in temporal integration and/or temporal processing. The evidence suggested that multiple individual stimuli can eventually be bound to a single memory trace in both modalities. However, a major difference between modalities seemed to be that in vision, stimulus order is lost when integration occurs, for which no clear auditory evidence had previously existed (cf. Akyürek et al., 2012; Näätänen and Winkler, 1999). Such a difference might have a clear functional origin, since hearing depends more on resolving temporal properties, if only because sound arrives in the ear in a sequential manner.

To provide an answer to this research question whether stimulus order is also lost in audition, we conducted a new experiment that allowed us to conclude that merging temporal integration in audition indeed behaves similar to how it behaves in vision. Using a task that is similar to the rapid serial visual presentation (RSVP) task - which can be used to investigate a merging form of temporal integration (i.e., in which two stimuli are integrated and combined into a single representation, and in which information about stimulus order is missing) - namely the rapid serial auditory presentation (RSAP) task, we found that sequential auditory stimuli can also be temporally integrated in the merging form. This entails that temporal integration is likely to be an amodal process, inherent to higher processes of perception, which consequently operates in a similar manner on information coming in from both the visual and auditory modalities.

These results tally with the amodal concept of the functional moment, which is defined as an interval in which temporal relations between stimuli are missing (Wittman, 2011; Dorato and Wittmann, 2015; Wittman, 2016). However, the results are not in line with an auditory theory of temporal integration that implies that stimuli are placed on temporal coordinates within a temporal integration window (Näätänen and Winkler, 1999), nor with theories that assume that no ‘long-range’ temporal integration takes place in audition (Viemeister, 1996). Without such integration occurring, it is hard to see how the dual tones we used as our target stimuli in the RSAP task could have been integrated across the temporal delays we used to form artificial vowels—the perception of which would normally require simultaneous presentation.

The temporal coordinates theory (Näätänen and Winkler, 1999) is more in line with auditory streaming phenomena, as herein successive stimuli are not combined into a single percept but rather regarded as a single stream (Bregman, 1994). This theory is mainly based on Mismatch Negativity (MMN) studies (e.g., Yabe et al., 1998; Sussman et al., 1999), which actually resemble our RSAP task more than auditory streaming tasks. Namely, during common MMN experiments, mostly one single auditory stimulus is repeated and occasionally one or two deviant auditory stimuli are presented, which elicit the MMN response. This setup is thus somewhat similar to our RSAP task, in which mostly one auditory distractor is repeated and occasionally one or two targets are presented. In these MMN studies, it was not tested whether participants heard one or two stimuli in cases when temporal integration occurred. However, in our RSAP task, participants reported perceiving a fair number of temporally integrated stimuli in settings that were optimal for temporal integration (i.e., at Lag 1). Therefore, because of the similarities in these tasks and the fact that participants did report perceiving temporally integrated stimuli in our RSAP task, we conclude that in the merging form of temporal integration, temporal coordinates of different stimuli in a temporal integration window are not retained but are rather missing, as is in line with the concept of a functional moment. Further supporting evidence for this idea can also be gleaned from another MMN study, conducted by Tervaniemi et al. (1994), as they did not find a significant MMN effect when the order of two deviants were reversed compared to when the order was not reversed. This also indicates that the perceptual system in this case did not observe the temporal order difference, which supports our conclusion.

In the third chapter, we focused on the effects of age on merging temporal integration. The question that we sought to answer in this chapter was whether the temporal integration window becomes longer when we get older. Such an effect might be expected to occur in general, because as we age cognitive processes seem to slow down (Salthouse, 1996), which might affect integration also. As in the previous chapter, we were also interested in examining possible differences between visual and auditory modalities, but now also combined with age effects. This was driven by a survey of the literature that showed considerable evidence of prolonged temporal integration in advanced age for vision (e.g., Di Lollo, Arnett, & Kruk, 1982), but not for audition (e.g., Näätänen, Kujala, & Winkler, 2011). This discrepancy between the modalities raises the interesting possibility that cognitive slowing of temporal integration is modality-specific and might depend on the relative importance of temporal processing within the respective modalities. Because of the relative importance of temporal properties for processing auditory information, compared to the spatial dominance in vision (Kubovy, 1988), it is conceivable that temporal processing is spared from decline in audition. To draw a parallel to muscle strength; it might be that you either use it, or you lose it (cf. Schooler, 2007).

To investigate these questions, we used both an RSVP and an RSAP task with 3 stimulus durations (40, 70 and 100 ms), comparing a group of younger adult participants with a group of older adult participants. In the visual domain, we found that older individuals temporally integrate more overall and, most importantly, at longer stimulus durations. This indicates that older individuals have longer temporal integration windows, most likely due to the effects of cognitive slowing. In the auditory domain, we found that the older individuals also integrate more overall, but only marginally more at longer stimulus durations, which indicates that the effects of cognitive slowing might indeed be slightly attenuated by the primacy of time in audition. This is in line with other studies on perceptual information processing in time. For example, gap detection thresholds and minimum inter stimulus interval (ISI) to judge temporal order are shorter for auditory stimuli than visual stimuli (Kanabus et al., 2012). But more importantly, visual inspection of Fig. 2 of Humes et al. (2009) shows that there is a greater difference for the visual than the auditory domain in gap detection thresholds between young and older listeners.

Nonetheless, it must be acknowledged that the overall trend in both modalities was similar in the present experiments. This similarity echoes the findings from chapter 2, in which visual and auditory integration were found to be very similar for young adults. Does this mean that the auditory modality suffers a comparatively heavier blow due to the loss of temporal resolution—in the very dimension that is most critical for audition? We speculate that prolonged temporal integration in audition may also carry some benefits that might offset the losses. Evidence suggests that (longer) temporal integration reduces mental effort and saves working memory capacity (Akyürek et al., 2017; Wolff et al., 2015). Furthermore, longer perceptual intervals may also carry benefits downstream, for higher-level cognitive processes. One potential candidate is the phonemic restoration process involved in understanding noise-interrupted speech, where the speech segments that are missing due to interruptions are restored with help of cognitive/linguistic processes, and by which the intelligibility of the interrupted speech is enhanced (Bashford & Warren, 1979; Bashford, Riener, & Warren, 1992; Başkent, 2012; Benard, Mensink, & Başkent, 2014). During phonemic restoration, having a longer integration window may afford the benefit of being able to use more contextual information, and may thereby be easier or more successful. The next chapter focused specifically on phonemic restoration and ageing.

In the fourth chapter we asked whether older individuals retain the ability to restore and understand degraded speech, which may be related to changes in temporal integration with age. In the discussion section of the second chapter we already discussed the potential relationship between the merging form of temporal integration and phonemic restoration. We concluded there that this merging form of temporal integration is most likely not related to or involved in phonemic restoration, mostly because speech segments in an integration window are not combined into a unitary percept. However, we hypothesized that a less merging form of temporal integration might still be involved in phonemic restoration.

To wit, we know that temporal integration is linked to functional moments, which are themselves merged into experienced moments that enable us to experience fluent perception (Wittmann, 2016). One possibility is that a less merging form of temporal integration is used to integrate segments of the speech signal, which are perceived by glimpsing, and which fills in the missing pieces of speech by using our vocabulary and language knowledge on contextual and sentential cues (Verschuure & Brocaar, 1983; Warren & Sherman, 1974). As an example, consider the bottom panels of Fig 4.3, where it is shown that most phonemic restoration benefit is obtained at 1.25 and 2.5 Hz, which respectively entail speech and noisy segments of 100 and 200 ms each. Note that to integrate and connect these speech segments requires spanning 200 and 400 ms, respectively. This timing might already pose something of a challenge for temporal integration. Apart from that, a form of temporal integration is probably not suitable for the purpose of this task to begin with, as the speech segments of the durations that are perceived in the glimpses should not be perceived as single, combined utterances, but instead as complex and dynamic units comprising syllables. For example, see the bottom middle panel of fig. 4.3, where most phonemic restoration is at 2.5 Hz, which is a frequency that lies within the syllabic rate of 2-5 Hz (Verhoeven et al., 2004; Edwards and Chang, 2013). In other words, the order of speech cues within a segment is mostly intact.

It is therefore more likely that speech comprehension is facilitated by a less merging form of temporal integration that connects functional moments of shorter duration or speech segments into experienced moments. This entails that this process involves top-down mechanisms that use language knowledge and vocabulary on contextual and sentential cues, and which uses Gestalt principles to infer what speech segments were missing during the segments without audible speech (Clarke, Gaudrain, Chatterjee, & Başkent, 2014). At the same time, the speech segments that are actually present have to be connected to the restored speech segments to form one continuous speech signal. This idea is in line with the asymmetric sampling in time hypothesis proposed by Poeppel (2003), which states that auditory information is analyzed in the brain asymmetrically. Namely, fast, spectral information (i.e., temporal fine structure) is analyzed in a temporal integration window of about 20-40 ms, while syllabic, prosodic and slow-envelope information are analyzed in a longer temporal integration window of about 150-250 ms in the right hemisphere. This entails the use of two different forms of temporal integration in speech perception; a functional moment of sorts to extract and analyze fast auditory information, and an experienced moment for slow information.

In our experiment, we compared phonemic restoration in groups of young and older healthy adult listeners. Advanced age is usually strongly associated with cognitive changes, which was the main interest of the study, as well as hearing loss, which was not. Previous studies with older individuals had not been controlled for hearing loss, and as a result, the effects reported were a combination of age-related cognitive change and hearing loss (Başkent, 2010; Başkent, Eiler, & Edwards, 2010). These earlier studies had indicated that older individuals with moderate hearing loss seem not to be able to benefit from phonemic restoration, while the ones with mild hearing loss seem to. For a more definitive answer, in the study reported in this thesis, hearing loss was controlled for by careful selection of older individuals with minimal amount of hearing loss. With this design, we found that the older group benefited equally from phonemic restoration compared to the younger group, and further, in a few situations they benefitted even more. For example, the bottom, middle panel of Fig 4.3 shows that older individuals had significantly more restoration benefit than the younger individuals at especially 2.5 Hz interruption rate. This can be explained by looking at the top, middle panel of the same figure, which shows that the older group had significantly lower intelligibility for speech interrupted with silence. As was noted before, 2.5 Hz falls within the 2 to 5 Hz syllabic rate, meaning that mostly syllables were erased. This introduces ambiguity in the speech signal and clears out a substantial number of sentential and contextual cues (Bhargava, Gaudrain and Başkent, 2016). However, the older individuals seem to be able to use their vocabulary and language skills better when the addition of noise to the silent parts activates the phonemic restoration mechanism. With slower interruption rates, larger parts of words or even entire words are alternately removed from or kept within the speech signal, giving more context to infer what was being said in the silent parts. With faster interruption rates, smaller segments of words and syllables are available, which makes it easier to perceive most of the words. These two situations, hence, seem to give less room for improvement when noise is added.

Interestingly, the lowest intelligibility for both age groups with speech interrupted with silence for slow and fast speech (see top left and top right panels of Fig 4.3) were at respectively 1.25 Hz and 5 Hz. The slow and fast speech rates were either slowed down or sped up by a factor of 2 with respect to the original speech rate. This manipulation was included to further probe into temporal processing in young and older individuals, in the context of phonemic restoration. The dips in intelligibility for speech interrupted with silence for the different speech rates also differed by a factor of 2 (i.e., 1.25 vs. 2.5 vs. 5 Hz.). Thus, it can be concluded that a change in speech rate shifts the syllabic rate and entails a proportional change in interruption rate to maintain the same pattern of intelligibility. This means that speech that is interrupted with silence at the syllabic rate should result in the lowest intelligibility. This effect indeed existed for both age groups, but seemed stronger for the older group as they had deeper dips in intelligibility. These syllabic rates, however, also enable them to gain more intelligibility when noise is added, or in other words, there is more room for the phonemic restoration mechanism to work on the ambiguity, allowing the older individuals to use their linguistic skills better. For the fast speech, the lowest interruption rates (i.e., 1.25, 2.5 and 5 Hz) did not show any restoration benefit, and there was even a large negative effect for the highest interruption rates (i.e., 10 and 20 Hz). Even though there was a large dip in intelligibility of speech interrupted with silence at 5 Hz, the noise did not improve intelligibility. Even more, at 10 and 20 Hz, intelligibility of speech interrupted with silence was at least as good as at 1.25 and 2.5 Hz. However, the addition of noise had a large negative effect on intelligibility for just the 10 and 20 Hz, most likely because the noise predominated the speech. The right middle panel of Fig 4.3 shows that the older group was significantly more affected by these extreme conditions than the younger group, most likely because they could not keep up with the speech due to cognitive slowing. This is evidence in line with the idea that the older group had larger temporal integration windows, which resulted in negative effects in these extreme listening conditions. Namely, signal information over longer periods was being analyzed, which then only meant that the noise dominated even more.

CONCLUSION

In general, the outcomes of the experiments in this thesis showed two things above all: First, comparing temporal integration across two modes of perception, namely vision and audition, showed that temporal processing in both shows great similarity. Second, exploring age effects on temporal integration showed that temporal processing in young and older adults is also very similar. On top of that, however, the thesis also provides evidence for differences, such as in the extent to which temporal integration covers longer intervals with ageing, and in the relative restoration rate of masked, inaudible speech segments. The overall picture is suggestive of a certain universality in how we process brief sensory input, and of graceful, sometimes even silver-lined, degradation, as we get older.

REFERENCES

Abate, M., Di Iorio, A., Di Renzo, D., Paganelli, R., Saggini, R., & Abate, G. (2007). Frailty in the elderly: The physical dimension. Europa Medicophysica, 43(3), 407–415.

Adams EM, Gordon-Hickey S, Morlas H, Moore R (2012) Effect of Rate-Alteration on Speech Perception in Noise in Older Adults With Normal Hearing and Hearing Impairment. Am J Audiol 21, 22–32. doi: 10.1044/1059-0889(2011/10-0023).

Akyürek, E. G., Eshuis, S. A. H., Nieuwenstein, M. R., Saija, J. D., Başkent, D., & Hommel, B. (2012). Temporal target integration underlies performance at lag 1 in the attentional blink. Journal of Experimental Psychology. Human Perception and Performance, 38(6), 1448–1464.

Akyürek, E. G., Kappelmann, N., Volkert, M., & van Rijn, H. (2017). What you see is what you remember: Visual chunking by temporal integration enhances working memory. Journal of Cognitive Neuroscience, 29, 2025-2036.

Akyürek, E. G., Schubö, A., & Hommel, B. (2010). Fast temporal event integration in the visual domain demonstrated by event-related potentials. Psychophysiology, 47(3), 512–522.

Akyürek, E. G., Toffanin, P., & Hommel, B. (2008). Adaptive control of event integration. Journal of Experimental Psychology. Human Perception and Performance, 34(3), 569–577.

Anderson S, White-Schwoch T, Parbery-Clark A, Kraus N (2013) Reversal of age-related neural timing delays with training. Proc Natl Acad Sci USA. doi: 10.1073/pnas.1213555110.

Baltes PB (1993) The Aging Mind: Potential and Limits. The Gerontologist, 33, 580–594. doi: 10.1093/geront/33.5.580.

Bashford, J. A., Riener, K. R., & Warren, R. M. (1992). Increasing the intelligibility of speech through multiple phonemic restorations. Perception & Psychophysics, 51, 211-217.

Bashford, J. A., & Warren, R. M. (1979). Perceptual synthesis of deleted phonemes. The Journal of the Acoustical Society of America, 65, S112.

Başkent, D. (2010). Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores. The Journal of the Acoustical Society of America, 128, EL169-EL174.

Başkent, D. (2012). Effect of speech degradation on top-down repair: phonemic restoration with simulations of cochlear implants and combined electric- acoustic stimulation. Journal of the Association for Research in Otolaryngology, 13, 683-692.

Başkent, D., Eiler, C. L., & Edwards, B. (2009). Effects of envelope discontinuities on perceptual restoration of amplitude-compressed speech. The Journal of the Acoustical Society of America, 125(6), 3995–4005.

Başkent, D., Eiler, C.L., & Edwards, B. (2010). Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss. Hearing Research, 260, 54-62.

Benard MR, Başkent D (2013a) Perceptual learning of temporally interrupted and spectrally degraded speech. J Acoust Soc Am, 136(3), 1344 – 1351.

Benard MR, Başkent D (2013b) Perceptual learning of interrupted speech. PLoS ONE 8:e58149. doi: 10.1371/journal.pone.0058149.

Benard, M. R., Mensink, J.S., & Başkent, D. (2014). Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities. Journal of the Acoustical Society of America, 135, EL88-EL94.

Bergman M, Blumenfeld VG, Cascardo D, et al. (1976) Age-related decrement in hearing for speech: Sampling and longitudinal studies. J Gerontol, 31, 533–538. doi: 10.1093/geronj/31.5.533.

Bhargava P, Gaudrain E, Başkent D (2013) Top-Down Restoration of Speech in Cochlear-Implant Users. Hearing Research, 309, 113–123.

Bhargava, P., Gaudrain, E., & Başkent, D. (2016). The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners. Journal of the Association for Research in Otolaryngology, 17, 475-491.

Bloom, J. S., & Hynd, G. W. (2005). The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition? Neuropsychology Review, 15(2), 59–71. doi:10.1007/s11065-005-6252-y.

Boersma P (2002) Praat, a system for doing phonetics by computer. Glot Int, 5, 341–345.

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. Psychological Review, 114(1), 38–70. doi:10.1037/0033-295X.114.1.38.

Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10(4), 433–436.

Bregman, A. S. (1994). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: MIT Press.

Broadbent, D. E., & Broadbent, M. H. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. Perception & Psychophysics, 42(2), 105–113.

Buckner RL (2004) Memory and executive function in aging and AD: Multiple factors that cause decline and reserve factors that compensate. Neuron, 44, 195–208.

Carlyon, R. P., Deeks, J., Norris, D., & Butterfield, S. (2002). The Continuity Illusion and Vowel Identification. Acta Acustica united with Acustica, 88(3), 408–415.

Cattell RB (1971) Abilities: their structure, growth, and action. xxii, 583.

Chanmugam, A., Triplett, P., & Kelen, G. (2013). Emergency Psychiatry. Cambridge University Press.

Charlton RA, Barrick TR, Markus HS, Morris RG (2010) The relationship between episodic long-term memory and white matter integrity in normal aging. Neuropsychologia, 48, 114–122. doi: 10.1016/j.neuropsychologia.2009.08.018.

Chen, T., & Naveh-Benjamin, M. (2012). Assessing the associative deficit of older adults in long-term and short-term/working memory. Psychology and Aging, 27(3), 666–682. doi:10.1037/a0026943.

Chua, F. K. (2005). The effect of target contrast on the attentional blink. Perception & Psychophysics, 67(5), 770–788.

Ciocca, V., & Bregman, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. Perception & Psychophysics, 42(5), 476–484.

Ciocca, V., & Darwin, C. J. (1999). The integration of nonsimultaneous frequency components into a single virtual pitch. The Journal of the Acoustical Society of America, 105(4), 2421–2430.

Clarke, J., Gaudrain, E., Chatterjee, M., & Başkent, D. (2014). T'ain't the way you say it, it's what you say – Perceptual voice continuity and top-down restoration of speech. Hearing Research, 325, 80-87.

Coltheart, M. (1980). Iconic memory and visible persistence. Perception & Psychophysics, 27, 183–228. doi: 10.3758/BF03204258.

Craik, F. I., & Salthouse, T. A. (2011). The handbook of aging and cognition. Psychology Press.

Dalton DS, Cruickshanks KJ, Klein BEK, et al. (2003) The impact of hearing loss on quality of life in older adults. Gerontologist, 43, 661–668. doi: 10.1093/geront/43.5.661.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. Hearing Research, 229(1-2), 132–147.

Di Lollo, V. (1980). Temporal integration in visual memory. Journal of Experimental Psychology: General, 109(1), 75–97. doi:10.1037/0096-3445.109.1.75.

Di Lollo, V., Arnett, J. L., & Kruk, R. V. (1982). Age-related changes in rate of visual information processing. Journal of Experimental Psychology: Human Perception and Performance, 8, 225–237.

Di Lollo, V., Hogben, J. H., Dixon, P. (1994). Temporal integration and segregation of brief visual-stimuli – Patterns of correlation in time. Perception & Psychophysics, 55, 373–386.

Dorato, M., & Wittmann, M. (2015). The now and the passage of time from physics to psychology. Kronoscope-Journal for the Study of Time, 15, 191-213.

Eddins, D. A., & Green, D. M. (1995). Temporal integration and temporal resolution. Hearing, 207–242.

Edwards, E., & Chang, E. F. (2013). Syllabic (similar to 2-5 Hz) and fluctuation (similar to 1-10 Hz) ranges in speech and auditory processing. Hearing Research, 305, 113-134.

Fogerty, D., Humes, L. E., & Busey, T. A. (2016). Age-related declines in early sensory memory: Identification of rapid auditory and visual stimulus sequences. Frontiers in Aging Neuroscience, 8. doi: 10.3389/fnagi.2016.00090.

Füllgrabe C (2013) Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss. Am J Audiol, 22(2), 213.

Gazzaley A, Clapp W, Kelley J, et al. (2008) Age-related top-down suppression deficit in the early stages of cortical visual memory processing. Proc Natl Acad Sci USA, 105, 13122–13126.

Gehr, S. E., & Sommers, M. S. (1999). Age differences in backward masking. The Journal of the Acoustical Society of America, 106(5), 2793–2799. doi:10.1121/1.428104.

Getzmann S, Falkenstein M (2011) Understanding of spoken language under challenging listening conditions in younger and older listeners: A combined behavioral and electrophysiological study. Brain Res, 1415, 8–22.

Gordon-Salant S, Fitzgibbons PJ (1993) Temporal factors and speech recognition performance in young and elderly listeners. J Speech Hear Res, 36, 1276–1285.

Gordon-Salant, S., & Fitzgibbons, P. J. (2001). Sources of age-related recognition difficulty for time-compressed speech. Journal of Speech, Language, and Hearing Research, 44, 709–719. doi: 10.1044/1092-4388(2001/056).

Goy H, Pelletier M, Coletta M, Pichora-Fuller MK (2013) The effects of semantic context and the type and amount of acoustical distortion on lexical decision by younger and older adults. Journal of Speech, Language, and Hearing Research. doi: 10.1044/1092-4388(2013/12-0053).

Grady C (2012) The cognitive neuroscience of ageing. Nat Rev Neurosci, 13, 491–505.

Groppe DM, Choi M, Huang T, et al. (2010) The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. Brain Research, 1361, 54–66. doi: 10.1016/j.brainres.2010.09.003.

He N, Mills JH, Ahlstrom JB, Dubno JR (2008) Age-related differences in the temporal modulation transfer function with pure-tone carriers. J Acoust Soc Am, 124, 3841–3849. doi: 10.1121/1.2998779.

Hedden T, Gabrieli JD (2004) Insights into the ageing mind: A view from cognitive neuroscience. Nat Rev Neurosci, 5, 87–96.

Heinrich, A., Carlyon, R. P., Davis, M. H., & Johnsrude, I. S. (2008). Illusory vowels resulting from perceptual continuity: a functional magnetic resonance imaging study. Journal of Cognitive Neuroscience, 20(10), 1737–1752.

Hoffman HJ, Dobie RA, Ko C-W, et al. (2012) Hearing threshold levels at age 70 years (65–74 years) in the unscreened older adult population of the United States, 1959–1962 and 1999–2006. Hearing Res, 33, 437–440. doi: 10.1097/AUD.0b013e3182362790.

Hogben, J. H., & Lollo, V. di. (1974). Perceptual integration and perceptual segregation of brief visual stimuli. Vision Research, 14(11), 1059–1069.

Hommel, B., & Akyürek, E. G. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. The Quarterly Journal of Experimental Psychology Section A, 58(8), 1415–1433.

Horváth, J., & Burgyán, A. (2011). Distraction and the auditory attentional blink. Attention, Perception, & Psychophysics, 73(3), 695–701.

Horváth, J., Czigler, I., Winkler, I., & Teder-Sälejärvi, W. A. (2007). The temporal window of integration in elderly and young adults. Neurobiology of Aging, 28(6), 964–975. doi:10.1016/j.neurobiolaging.2006.05.002.

Hughes, J. W. (1946). The Threshold of Audition for Short Periods of Stimulation. Proceedings of the Royal Society of London. Series B, Biological Sciences, 133(873), 486–490.

Humes, L. E., Busey, T. A., Craig, J. C., & Kewley-Port, D. (2009). The effects of age on sensory thresholds and temporal gap detection in hearing, vision, and touch. Attention, Perception, & Psychophysics, 71, 860–871.

Humes LE, Lee JH, Coughlin MP (2006) Auditory measures of selective and divided attention in young and older adults using single-talker competition. J Acoust Soc Am, 120, 2926–2937. doi: 10.1121/1.2354070.

Humes, L. E., Busey, T. A., Craig, J., & Kewley-Port, D. (2013). Are age-related changes in cognitive function driven by age-related changes in sensory processing? Attention, Perception, & Psychophysics, 75(3), 508–524. doi: 10.3758/s13414-012-0406-9.

International Council of Ophthalmology : Resources : Visual Standards - Aspects And Ranges Of Vision Loss. (n.d.). Retrieved February 7, 2017, from http://www.icoph.org/resources/10/Visual-Standards---Aspects-and-Ranges-of-Vision-Loss.html.

ISO, B. (2003). 226: 2003:“Acoustics—Normal equal loudness-level contours”. International Organization for Standardization.

Janse E (2012) A non-auditory measure of interference predicts distraction by competing speech in older adults. Aging Neuropsychol C, 19, 741–758. doi: 10.1080/13825585.2011.652590.

Jeffress, L. A. (1967). Stimulus-Oriented Approach to Detection Re-Examined. The Journal of the Acoustical Society of America, 41(2), 480–488.

Kanabus, M., Szelag, E., Rojek, E., & Poppel, E. (2002). Temporal order judgement for auditory and visual stimuli. Acta Neurobiologiae Experimentalis, 62, 263-270.

Kemper S, Herman RE, T H (2003) The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech of noise. Psychol Aging, 18, 181–192. doi: 10.1037/0882-7974.18.2.181.

Kennedy KM, Raz N (2009) Aging white matter and cognition: Differential effects of regional variations in diffusion properties on memory, executive functions, and speed. Neuropsychologia, 47, 916–927. doi: 10.1016/j.neuropsychologia.2009.01.001.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. The Journal of the Acoustical Society of America, 67(3), 971–995. doi:10.1121/1.383940.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. The Journal of the Acoustical Society of America, 87(2), 820–857. doi:10.1121/1.398894.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., & others. (2007). What's new in Psychtoolbox-3. Perception, 36, 14–14.

Kline, D. W., Ikeda, D. M., & Schieber, F. J. (1982). Age and temporal resolution in color vision: When do red and green make yellow? Journal of Gerontology, 37(6), 705–709. doi:10.1093/geronj/37.6.705.

Kline, D. W., & Orme-Rogers, C. (1978). Examination of stimulus persistence as the basis for superior visual identification performance among older adults. Journal of Gerontology, 33(1), 76–81. doi:10.1093/geronj/33.1.76.

Kolodziejczyk, I., & Szelag, E. (2008). Auditory perception of temporal order in centenarians in comparison with young and elderly subjects. Acta Neurobiologiae Experimentalis, 68(3), 373–81.

Kubovy, M. (1988). Should we resist the seductiveness of the space: time: vision: Audition analogy. Journal of Experimental Psychology: Human Perception and Performance, 14, 318–320.

Lahar, C. J., Isaak, M. I., & McArthur, A. D. (2001). Age differences in the magnitude of the attentional blink. Aging, Neuropsychology, and Cognition, 8(2), 149–159. doi:10.1076/anec.8.2.149.842.

Lindenberger U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. Psychology and Aging, 9, 339–355. doi: 10.1037//0882-7974.9.3.339.

Lu, P. H., Lee, G. J., Raven, E. P., Tingus, K., Khoo, T., Thompson, P. M., & Bartzokis, G. (2011). Age-related slowing in cognitive processing speed is associated with myelin integrity in a very healthy elderly sample. Journal of Clinical and Experimental Neuropsychology, 33(10), 1059–1068. doi:10.1080/13803395.2011.595397.

Lu, P. H., Lee, G. J., Tishler, T. A., Meghpara, M., Thompson, P. M., & Bartzokis, G. (2013). Myelin breakdown mediates age-related slowing in cognitive processing speed in healthy elderly men. Brain and Cognition, 81(1), 131–138. doi:10.1016/j.bandc.2012.09.006.

Madden, D., & Allen, P. A. (2015). History of Cognitive Slowing Theory and Research. In N. A. Pachana (Ed.), Encyclopedia of Geropsychology, 1–10. Singapore: Springer.

Mahncke HW, Connor BB, Appelman J, et al. (2006) Memory enhancement in healthy older adults using a brain plasticity-based training program: A randomized, controlled study. Proc Natl Acad Sci USA 103:12523–12528. doi: 10.1073/pnas.0605194103.

Martini, A. (1996). European Working Group on Genetics of Hearing Impairment Infoletter 2, European Commission Directorate. Biomedical and Health Research Programme (HEAR), b21.

McCoy SL, Tun PA, Cox LC, et al. (2005) Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. Q J Exp Psychol A, 58, 22–33. doi: 10.1080/02724980443000151.

McDermott JH, Oxenham AJ (2008) Spectral completion of partially masked sounds. Proc Natl Acad Sci USA 105:5939–5944. doi: 10.1073/pnas.0711291105.

Micheyl, C., Carlyon, R. P., Shtyrov, Y., Hauk, O., Dodson, T., & Pullvermüller, F. (2003). The Neurophysiological Basis of the Auditory Continuity Illusion: A Mismatch Negativity Study. Journal of Cognitive Neuroscience, 15(5), 747–758.

Miller GA, Licklider JCR (1950) The Intelligibility of Interrupted Speech. J Acoust Soc Am, 22, 167. doi: 10.1121/1.1906584.

Moore, B. C. J. (2003). Temporal integration and context effects in hearing. Journal of Phonetics, 31(3), 563–574.

Moulines E, Charpentier F (1990) Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Commun, 9, 453–467. doi: 10.1016/0167-6393(90)90021-Z.

Munson, W. A. (1947). The Growth of Auditory Sensation. The Journal of the Acoustical Society of America, 19(4), 584–591.

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. Ear and Hearing, 16(1), 6–18.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. Behavioral and Brain Sciences, 13(2), 201–288.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clinical Neurophysiology, 118(12), 2544–2590.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. Psychological Bulletin, 125(6), 826–859.

Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. Psychophysiology, 48(1), 4–22.

O'Callaghan, C. (2008). Object perception: Vision and audition. Philosophy Compass, 3(4), 803–829.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. Biometrics, 57, 120–125. doi: 10.1111/j.0006-341X.2001.00120.x.

Park DC, Lautenschlager G, Hedden T, et al. (2002) Models of visuospatial and verbal memory across the adult life span. Psychol Aging, 17, 299–320. doi: 10.1037/0882-7974.17.2.299

Pedersen, C. B., & Elberling, C. (1972). Temporal integration of acoustic energy in normal hearing persons. Acta Oto-Laryngologica, 74(6), 398–405.

Pedersen, C. B., & Salomon, G. (1977). Temporal integration of acoustic energy. Acta Oto-Laryngologica, 83(5-6), 417–423.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10(4), 437–442.

Pichora-Fuller MK (2008) Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. Int J Audiol, 47, S72–S82. doi: 10.1080/14992020802307404.

Plomp, R., & Bouman, M. A. (1959). Relation between Hearing Threshold and Duration for Tone Pulses. The Journal of the Acoustical Society of America, 31(6), 749–758.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication, 41, 245–255.

Pols, L. C. W., Tromp, H. R. C., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. The Journal of the Acoustical Society of America, 53(4), 1093–1101.

Powers GL, Wilcox JC (1977) Intelligibility of temporally interrupted speech with and without intervening noise. J Acoust Soc Am, 61, 195–199. doi: 10.1121/1.381255.

Rabiner, L. R., & Schafer, R. W. (1978). Digital processing of speech signals (Vol. 100). Prentice-hall Englewood Cliffs, NJ.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? Journal of Experimental Psychology: Human Perception and Performance, 18(3), 849–860.

Reuter-Lorenz PA (2002) New visions of the aging mind and brain. Trends Cogn Sci, 6, 394–400.

Riecke L, Mendelsohn D, Schreiner C, Formisano E (2009) The continuity illusion adapts to the auditory scene. Hearing Res, 247, 71–77. doi: 10.1016/j.heares.2008.10.006.

Riecke, L., Opstal, A. J. van, & Formisano, E. (2008). The auditory continuity illusion: A parametric investigation and filter model. Perception & Psychophysics, 70(1), 1–12.

Roberts, K. L., & Allen, H. A. (2016). Perception and cognition in the ageing brain: A brief review of the short- and long-term links between perceptual and cognitive decline. Frontiers in Aging Neuroscience, 8. doi: 10.3389/fnagi.2016.00039.

Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. Nature, 398(6730), 760–760.

Saija, J. D., Andringa, T. C., Başkent, D., & Akyürek, E. G. (2014a). Temporal integration of consecutive tones into synthetic vowels demonstrates perceptual assembly in audition. Journal of Experimental Psychology: Human Perception and Performance, 40(2), 857–869. doi: 10.1037/a0035146.

Saija, J. D., Akyürek, E. G., Andringa, T. C., & Başkent, D. (2014b). Perceptual restoration of degraded speech is preserved with advancing age. Journal of the Association for Research in Otolaryngology, 15, 139-148. doi: 10.1037/a0035146.

Salami, A., Eriksson, J., Nilsson, L.-G., & Nyberg, L. (2012). Age-related white matter microstructural differences partly mediate age-related decline in processing speed but not cognition. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1822(3), 408–415. doi:10.1016/j.bbadis.2011.09.001.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. Psychological Review, 103, 403–428.

Salthouse TA (2004) What and when of cognitive aging. Curr Dir Psychol Sci, 13, 140–144. doi: 10.1111/j.0963-7214.2004.00293.x.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. Journal of Experimental Psychology: General, 110(4), 474–494.

Schneider, B. A., Daneman, M., & Murphy, D. R. (2005). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? Psychology and Aging, 20, 261–271. doi: 10.1037/0882-7974.20.2.261.

Schooler, C. (2007). Use it—and keep it, longer, probably: A reply to Salthouse (2006). Perspectives on Psychological Science, 2, 24–29.

Shahin AJ, Bishop CW, Miller LM (2009) Neural mechanisms for illusory filling-in of degraded speech. NeuroImage, 44, 1133–1143. doi: 10.1016/j.neuroimage.2008.09.045.

Sheldon S, Pichora-Fuller MK, Schneider BA (2008) Priming and sentence context support listening to noise-vocoded speech by younger and older adults. J Acoust Soc Am, 123, 489–499. doi: 10.1121/1.2783762.

Slawinski, E. B., & Goddard, K. M. (2001). Age-related changes in perception of tones within a stream of auditory stimuli: Auditory attentional blink. Canadian Acoustics, 29(1), 3–12.

Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior knowledge during speech perception. J Neurosci, 32, 8443–8453.

Srinivasan S, Wang D (2005) A schema-based model for phonemic restoration. Speech Commun, 45, 63–87.

Stenfelt, S., & Rönnberg, J. (2009). The Signal-Cognition interface: Interactions between degraded auditory signals and cognitive processes. Scandinavian Journal of Psychology, 50(5), 385–393.

Stephens D (1996) Hearing rehabilitation in a psychosocial framework. Scand Audiol Suppl, 43, 57–66.

Swenor BK, Ramulu PY, Willis JR, et al. (2013) The prevalence of concurrent hearing and vision impairment in the United States. JAMA Intern Med, 173, 312–313. doi: 10.1001/jamainternmed.2013.1880.

Sussman, E., Winkler, I., Ritter, W., Alho, K., & Näätänen, R. (1999). Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity. Neuroscience Letters, 264, 161–164.

Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., & Näätänen, R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. Biological Psychology, 38, 157–167.

Tonoki A, Davis RL (2012) Aging impairs intermediate-term behavioral memory by disrupting the dorsal paired medial neuron memory trace. Proc Natl Acad Sci USA, 109, 6319–6324. doi: 10.1073/pnas.1118126109.

Treisman, A. (1996). The binding problem. Current Opinion in Neurobiology, 6(2), 171–178.

Tremblay, S., Vachon, F., & Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. Perception & Psychophysics, 67(2), 195–208.

Ulbrich, P., Churan, J., Fink, M., & Wittmann, M. (2009). Perception of temporal order: The effects of age, sex, and cognitive factors. Aging, Neuropsychology, and Cognition, 16(2), 183–202. doi: 10.1080/1382558080241175.

Verschuure J, Brocaar MP (1983) Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. Percept Psychophys, 33, 232–240. doi: 10.3758/BF03202859.

Versfeld NJ, Daalder L, Festen JM, Houtgast T (2000) Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J Acoust Soc Am, 107, 1671–1684. doi: 10.1121/1.428451.

Viemeister, N. (1996). Auditory temporal integration: What is being accumulated? Current Directions in Psychological Science, 5, 28–32.

Viemeister, N. F., & Wakefield, G. H. (1991). Temporal integration and multiple looks. The Journal of the Acoustical Society of America, 90(2), 858–865.

Visser, T. A. W., & Enns, J. T. (2001). The role of attention in temporal integration. Perception, 30, 135–145. doi: 10.1068/p3089.

Verhoeven, J., De Pauw, G., & Kloots, H. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. Language and Speech, 47, 297-308.

Verschuure, J., & Brocaar, M. P. (1983). Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. Perception & Psychophysics, 33, 232-240.

Wallace, A. B., & Blumstein, S. E. (2009). Temporal integration in vowel perception. Journal of the Acoustical Society of America, 125, 1704–1711. doi: 10.1121/1.3077219.

Warren RM (1970) Perceptual restoration of missing speech sounds. Science, 167, 392–393.

Warren, R. M. (1999). Auditory perception: A new analysis and synthesis (Vol. xiv). New York, NY, US: Cambridge University Press.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. Science, 176(4039), 1149–1151.

Warren, R. M., & Sherman, G. L. (1974). Phonemic restorations based on subsequent context. Perception & Psychophysics, 16, 150-156.

Wayne, R. V., & Johnsrude, I. S. (2015). A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. Ageing Research Reviews, 23, 154–166. doi: 10.1016/j.arr.2015.06.002.

Weale, R. A. (1963). New light on old eyes. Nature, 198, 944–946. doi: 10.1038/198944b0.

Weenink, D., & others. (2009). The klattgrid speech synthesizer. In Interspeech, 2059–2062.

Wild-Wall N, Falkenstein M (2010) Age-dependent impairment of auditory processing under spatially focused and divided attention: An electrophysiological study. Biol Psychol, 83, 27–36. doi: 10.1016/j.biopsycho.2009.09.011.

Willems, C., Saija, J. D., Akyürek, E. G., & Martens, S. (2016). An Individual Differences Approach to Temporal Integration and Order Reversals in the Attentional Blink Task. PLoS ONE, 11, e0156538. doi: 10.1371/journal.pone.0156538.

Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. Journal of the American Academy of Audiology, 7, 175–182.

Wingfield A, Tun PA, McCoy SL (2005) Hearing loss in older adulthood what it is and how it interacts with cognitive performance. Curr Dir Psychol Sci, 14, 144–148. doi: 10.1111/j.0963-7214.2005.00356.x.

Wittmann, M. (2011). Moments in time. Frontiers in Integrative Neuroscience, 5.

Wittmann, M. (2016). The duration of presence. In B. Mölder, V. Arstila, & P. Øhrstrøm (Eds.), Philosophy and Psychology of Time, 101–113. Springer International Publishing.

Wolff, M. J., Scholz, S., Akyürek, E. G., & van Rijn, H. (2015). Two visual targets for the price of one? Pupil dilation shows reduced mental effort through temporal integration. Psychonomic Bulletin & Review, 22, 251–257.

Wong PCM, Jin JX, Gunasekera GM, et al. (2009) Aging and cortical mechanisms of speech perception in noise. Neuropsychologia, 47, 693–703. doi: 10.1016/j.neuropsychologia.2008.11.032.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. Psychophysiology, 35, 615–619.

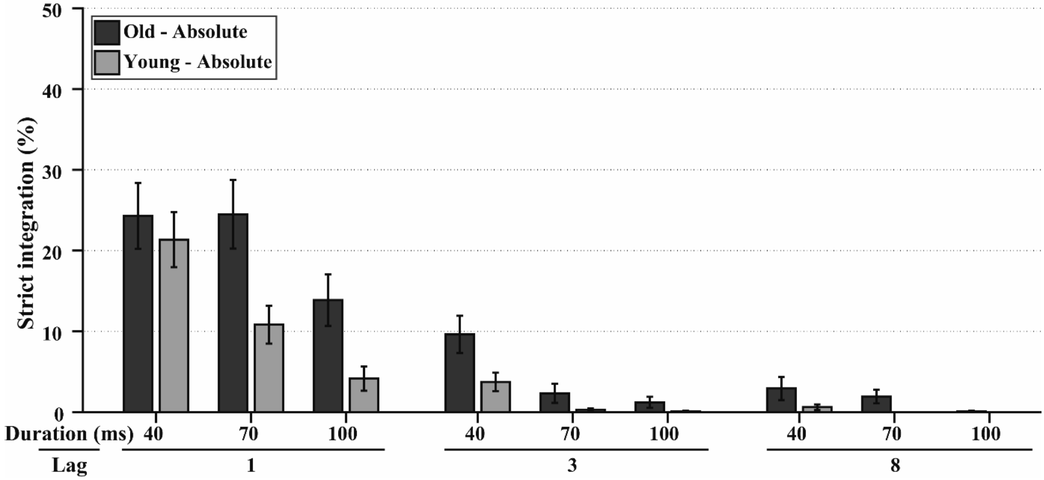
Yu, L., Yabe, H., Shiga, T., Nozaki, M., Ohshima, H., Itagaki, S., … Niwa, S. (2011). Only a stimulus onset might initiate the temporal window of integration. In 2011 IEEE/ICME International Conference on Complex Medical Engineering (CME), 246 –247. Presented at the 2011 IEEE/ICME International Conference on Complex Medical Engineering (CME).

Zwislocki, J. J. (1960). Theory of Temporal Auditory Summation. The Journal of the Acoustical Society of America, 32(8), 1046–1060.

Zwislocki, J. J. (1969). Temporal Summation of Loudness: An Analysis. The Journal of the Acoustical Society of America, 46(2B), 431–441.

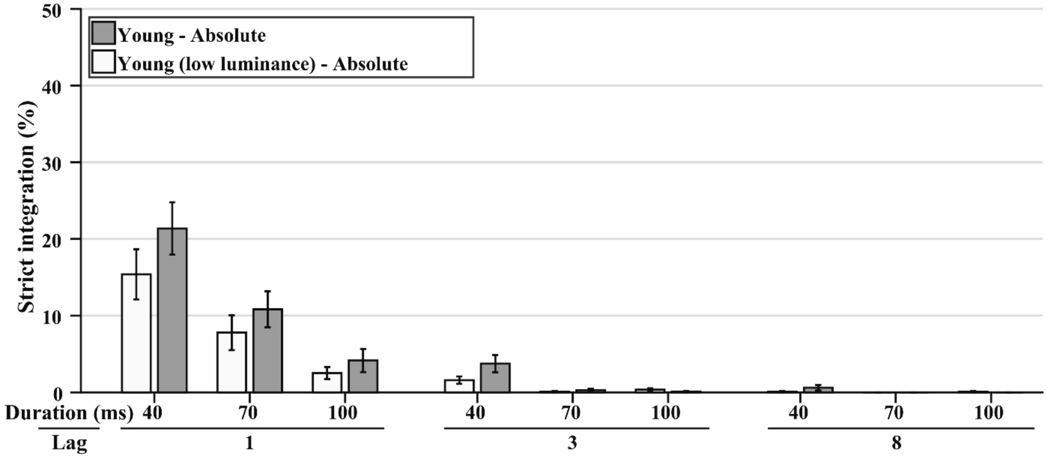
APPENDIX

APPENDIX FIGURE 1



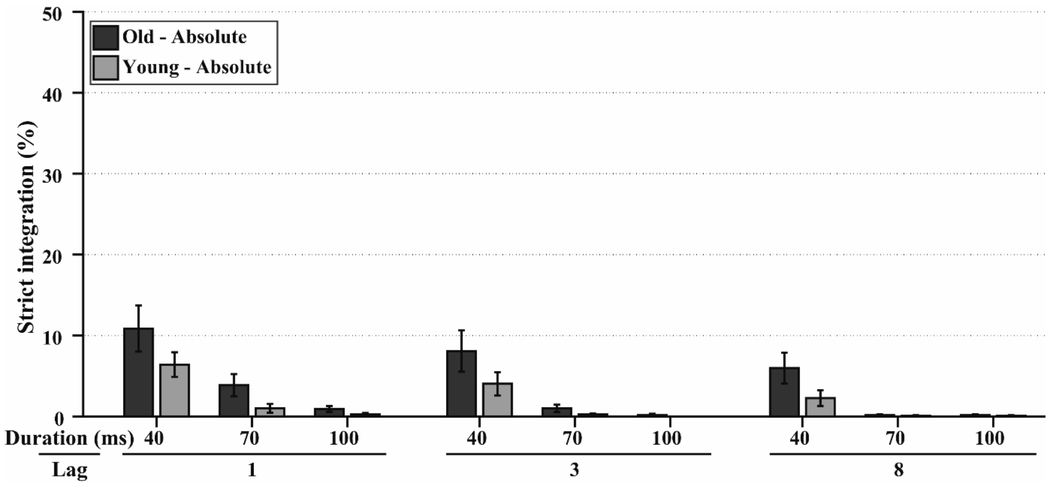
Appendix Figure 1 - Experiment 1A: Estimated marginal means of the analyses of absolute frequency of strict visual integrations for all combinations of stimulus duration, lag and age group, as a percentage of the total number of trials in which both target identities were preserved. Error bars represent ± 1 standard error of the mean.

APPENDIX FIGURE 2



Appendix Figure 2 - Experiment 1B: Estimated marginal means of the analyses of absolute frequency of strict visual integrations for all combinations of stimulus duration and lag, as a percentage of the total number of trials in which both target identities were preserved. The data from the young group of Experiment 1A (full luminance) are re-plotted next to the low luminance group for reference. Error bars represent ± 1 standard error of the mean.

APPENDIX FIGURE 3



Appendix Figure 3 - Experiment 2: Estimated marginal means of the analyses of absolute frequency of strict auditory integrations for all combinations of stimulus duration, lag and age group, as a percentage of the total number of trials in which both target identities were preserved. Error bars represent ± 1 standard error of the mean.

NEDERLANDSE SAMENVATTING

In dit proefschrift zijn verschillende aspecten van het temporele integratieproces onderzocht, een perceptueel en cognitief proces waardoor snel na elkaar binnenkomende stukjes informatie worden gegroepeerd en in een enkele omvattende 'gebeurtenis' (event) worden geplaatst. Het hier gerapporteerde onderzoek richt zich daarbij vooral op mogelijke overeenkomsten tussen perceptuele informatie die via de visuele en via de auditieve waarneming wordt verwerkt. Daarbij hadden ook mogelijke vertragende effecten van cognitieve veroudering op de verwerking in deze modaliteiten onze specifieke aandacht. Concreet stelden we de volgende onderzoeksvragen in de empirische hoofdstukken van dit proefschrift:

1. Is temporele integratie vergelijkbaar in de visuele en auditieve modaliteit (Hoofdstuk 2)?
2. Wordt temporele integratie vertraagd door cognitieve veroudering, en is dit vergelijkbaar in de visuele en auditieve modaliteit (Hoofdstuk 3)?
3. Kan er door hogere cognitieve functies ook een voordeel gehaald worden door vertraagde perceptuele verwerking, gerelateerd aan temporele integratie (Hoofdstuk 4)?

De eerste onderzoeksvraag kwam voort uit een lezing van de literatuur over temporele integratie in de visuele en auditieve modaliteiten. Hieruit bleek dat er in beide modaliteiten inderdaad bewijs is dat meerdere achtereenvolgende stimuli in een enkele gebeurtenis worden verwerkt. Het bleek echter ook dat er in de auditieve modaliteit geen bewijs was voor het verlies van volgordelijke informatie over de individuele stimuli, iets wat in de visuele modaliteit juist evident is. Omdat het gehoor sterk rust op het interpreteren van snel opeenvolgende klanken (zoals bijvoorbeeld bij het begrijpen van spraak), zou dit een functioneel onderscheid tussen de modaliteiten kunnen zijn.

Om de mogelijke overeenkomsten tussen de visuele en auditieve modaliteiten te onderzoeken, maakten wij gebruik van een auditieve experimentele taak die voor zover mogelijk gelijkwaardig was aan een bestaande visuele taak. Dit betrof de zogenaamde Rapid Serial Auditory Presentation (RSAP) taak, welke dus gebaseerd was op de Rapid Serial Visual Presentation (RSVP) taak. In deze taken worden stimuli snel achter elkaar aangeboden, elk voor een duur van circa 100 msec in een stroom van ongeveer 20 items. In deze stroom bevinden zich twee doelstimuli (targets), die op volgorde geïdentificeerd moeten worden door de proefpersoon. De mate waarin zij hierin slagen hangt af van de tijd tussen de targets in de stroom: bij kortere intervallen speelt een aandachtsprobleem (de zogenaamde Attentional Blink; AB), en bij de meest korte intervallen kunnen – indien mogelijk – de twee targets temporeel geïntegreerd worden. In de RSAP hield dit in dat twee opeenvolgende tonen waargenomen zouden moeten worden als de simultane combinatie daarvan, resulterend in de waarneming van een kunstmatige klinker. De resultaten wezen uit dat de auditieve targets inderdaad op deze wijze geïntegreerd werden in de RSAP taak, wat vergelijkbaar is met de resultaten van eerdere RSVP-studies. Dit betekent dus dat er waarschijnlijk sprake is van een a-modaal perceptueel moment; een tijdsinterval waarover successievelijke stimuli geïntegreerd worden, onafhankelijk van hun sensorische oorsprong. Deze bevinding is opvallend, omdat er theorieën van auditieve waarneming bestaan die deze vorm van integratie niet onderschrijven (e.g., Viemeister, 1996). De resultaten passen daarom beter bij alternatieve theorieën die wel rekening houden met integratie in deze vorm (e.g., Näätänen & Winkler, 1999).

De tweede onderzoeksvraag kwam voort uit de gedachte dat de vaak geobserveerde vertraging van cognitieve processen door veroudering (Salthouse, 1996) ook invloed zou kunnen hebben op temporele integratie, namelijk dat deze een langer tijdsinterval zou kunnen gaan omvatten. Uit lezing van de literatuur over visuele temporele verwerking bleek inderdaad dat hiervoor veel aanwijzingen waren. Met betrekking tot auditieve temporele verwerking was dit echter niet het geval. Deze discrepantie zou verklaard kunnen worden door een werkelijk, onderliggend verschil: Omdat het gehoor bij uitstek leunt op het verwerken van subtiele temporele eigenschappen van het binnenkomende geluid zou dit een meer primair aspect van de perceptuele verwerking kunnen zijn dan dat het geval is voor het gezichtsvermogen, waar met name spatiele aspecten primair zijn. Als de effecten door veroudering met name optreden voor functies die minder vaak gebruikt worden, net als dat bijvoorbeeld het geval is voor weinig gebruikte spieren, dan zou het kunnen dat temporele verwerking in de auditieve modaliteit minder vertraagd wordt dan in de visuele modaliteit.

Om dit te onderzoeken lieten we groepen jongere en oudere volwassenen deelnemen aan zowel een RSAP als een RSVP taak, waarin we bovendien de snelheid van de stimuli aanpasten van 40 tot 100 ms per stuk om mogelijke verandering over een zekere reikwijdte van het integratieproces te kunnen testen. In de visuele taak bleek dat de ouderen inderdaad meer integreren, met name als de stimuli langer duren. Dit bevestigt dat er inderdaad een vertragend effect van veroudering is op temporele integratie. In de auditieve taak integreerden de ouderen ook vaker, maar slechts marginaal meer met langer durende stimuli. Dit zou erop kunnen wijzen dat er toch een mate van behoud van functionaliteit is in het auditieve domein. Het moet wel worden benadrukt dat de algehele trend in beide modaliteiten vergelijkbaar was. We merkten hierbij op dat hoewel vooral de auditieve achteruitgang pijnlijk lijkt, vanwege het belang van temporele verwerking in deze modaliteit, het denkbaar is dat een langere integratie ook voordelen biedt. Eerder onderzoek liet zien dat door temporele integratie mentale inspanning en kostbare werkgeheugencapaciteit uitgespaard kan worden (Akyürek et al, 2017; Wolff et al., 2015). Vanuit dat licht bezien zou tragere integratie door veroudering zelfs als een positieve verandering kunnen worden beschouwd.

In het vierde hoofdstuk stuitten we op verder bewijs voor een positief effect van veroudering op een deels op temporele integratie gebaseerd proces, namelijk fonemische restauratie. Fonemische restauratie is het perceptuele herstel van spraak als die door andere geluiden deels wordt gemaskeerd. Dit herstel wordt mogelijk gemaakt door spraaksegmenten op te delen in gepaste stukken (i.e., een geheel woord), en dus ook de segmenten die van elkaar gescheiden zijn door andere, maskerende geluiden. Dat is eigenlijk een vorm van temporele integratie, maar fonemische restauratie berust daarbij ook op hogere cognitieve faciliteiten, met name kennis van de betreffende grammatische regels en de relevante vocabulaire.

We testten fonemische restauratie in een groep jongere en een groep oudere volwassenen, door onderbroken spraakfragmenten aan te bieden op verschillende snelheden, waarbij de onderbrekingen soms met een ander generiek geluid werden gevuld, wat restauratie mogelijk maakt. Omdat gehoorverlies veel voorkomt bij ouderen, en omdat dit de resultaten zou kunnen beïnvloeden, controleerden wij onze proefpersonen hier zorgvuldig op. Uit de resultaten bleek dat ouderen net als jongeren profiteren van de mogelijkheid om fonemische restauratie toe te passen (i.e., wanneer de onderbrekingen met een ander geluid gevuld waren). Dit effect werd nog sterker voor ouderen wanneer de spraak vertraagd werd afgespeeld, waardoor er meer tijd is om de spraak cognitief te verwerken. Deze uitkomsten steunen het idee dat temporele integratie ook hier een langer interval besloeg bij de oudere volwassenen, en dat dit in dit geval geen straf is, omdat het tot verbeteringen van het spraakbegrip kan leiden.

ACKNOWLEDGEMENTS

I would like to show my gratitude to my supervisors, Deniz, Elkan and Tjeerd for all the help, advice, good talks and fun we had. Thanks Deniz for showing me how academic writing is done! Also, thank you for your hospitality, for organizing a party at your place each year to start off the new academic year! Thanks Elkan for the nice conversations we had and for your guidance throughout the project! Thanks Tjeerd for helping me out where needed!

A big thank you to the reading and the defense committee!

Thanks to all people who were willing to participate in these experiments!

A thank you to my colleagues Aytaç, Enja, Nawal, Elouise, Daniël, Carina, Etienne and all other colleagues from the ENT- and psychology department! Thanks for the good times and the help!

I would also like to thank Paolo for the enjoyable conversations and all the help around the office and the lab! Also thanks to Ria and Ans for always helping and taking care of all the official work!

A special thanks to my office mates, Pranesh-San, Jeanne and Annika.. Thanks guys, for all the laughter, fun and good times we had! It was really a blast sharing the office with you guys. But also, thanks for the help where needed and the good talks we had. And Pranesh… Yoga fire!

And of course, much thanks to my family and friends!